Training the Compassionate and the Empathic Brain

Thesis presented to the Faculty of Arts of the University of Zurich for the degree of Doctor of Philosophy

by

Olga M Klimecki

of Poznan, Poland

Accepted in the fall semester 2012 on the recommendation of Prof. Dr. Lutz Jäncke and Prof. Dr. Tania Singer

Zurich 2012

[last pages Curriculum Vitae]

Acknowledgments

My deep gratitude goes to all the people who supported me throughout the inception, execution, and completion of my dissertation.

The first part of my PhD thesis was carried out from October 2007 to August 2010 with Prof. Dr. Tania Singer at the Laboratory for Social and Neural Systems Research at the University of Zurich, Switzerland. In September 2010, I followed Prof. Singer to the Max-Planck Institute for Human Cognitive and Brain Sciences in Leipzig, Germany.

First and foremost, my gratitude goes to Prof. Dr. Tania Singer for her scientific guidance and for inspiring me with her vision in the study of the neuronal and affective plasticity of compassion. During my time in her laboratory, I learned countless different skills pertaining to the planning and designing of cutting-edge experiments, applying for grants, conducting longitudinal intervention studies, and also analysis and publication of complex data sets. I very much appreciated the discussions we had, her methodological advice, and her detailed feedback throughout the writing process. Moreover, I am very grateful that, along with advancing in my scientific skills, I had the opportunity to grow personally on my spiritual path.

At both hosting institutions I was extremely fortunate to work with people who happily shared their scientific and practical expertise. I thank all my colleagues in Zurich and Leipzig for their support and my colleagues in Maastricht (Dr. Bettina Sorger in particular) for their co-operation. My gratitude especially goes to Dr. Susanne Leiberg for her scientific advice and the discussions we had throughout all phases of the studies, and for conducting the experiments with me. I thank Prof. Dr. Claus Lamm, Dr. Boris Bernhardt, and Dr. Nikolaus Steinbeis for their methodological advice and Dr. Matthieu Ricard for sharing his first person perspective on contemplative practices. Dr. Ricard, along with Catherine Felder, Ariya Nani, Jotika Hermsen, and many other spiritual teachers greatly helped to advance my understanding of Eastern contemplative practices related to compassion. By teaching the participants involved in the presented studies, Catherine Felder, Ariya Nani, and Jotika

Hermsen made it possible for me to investigate the plasticity of compassion and empathy. Annette Rentz-Lühning was so kind to fly in twice from Berlin in order to teach the Method of Loci in Zurich. I thank all the teachers involved for sharing their skills and wisdom so openly. The practical expertise from the people working in administration, data management, and graphics helped to provide the infrastructure for this project and to improve its presentation. Finally, I am deeply grateful to the research assistants and interns who helped to conduct the multitude of studies and to the volunteers who participated in these studies.

Lastly, my deep gratitude goes to Prof. Dr. Lutz Jäncke for kindly offering to head my PhD steering committee.

Financially, this work was supported by the University of Zurich ("Neural Foundations of Empathy Learning", Forschungskredit, grant number 53990601 to O.M.K.), grants to T.S. from the Neuroscience Center Zurich, the Swiss National Science Foundation (SNF NCCR Affective Sciences: "Neuronal and developmental mechanisms of human empathy"; SNF NCCR Neural Plasticity and Repair: "Plasticity of the empathic and attentional brain"), the Foundation for Research in Science and Humanities at the University of Zurich ("Neural foundations of empathy training"), and a Francisco J. Varela Research Award to S.L. from the Mind and Life Institute. The research leading to the results presented in this thesis has also received funding from the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013/ ERC Grant agreement number 205557 to T.S.).

Summary English

The main objective of the present thesis was to study plasticity in the social brain and to determine behavioral, affective, and neural changes induced by training compassion and empathy. In order to reliably measure longitudinal changes in this domain, two new tasks were developed and validated: the Zurich Prosocial Game (ZPG) as a novel implicit measure for helping behavior and the Socio-affective Video Task (SoVT) for measuring neural and affective responses to social stimuli.

In **Study 1**, the ZPG was validated with a sample of 68 female participants. Results show that the game is sensitive to the manipulation of reciprocity, distress cues, and helping cost. Using the ZPG in a short-term intervention study revealed that compassion (*n* = 27), but not memory training (*n* = 32) increased prosocial behavior in a training-unrelated task. Furthermore, in the compassion group, inter-individual differences in practice duration were parametrically related to changes in helping behavior under the no-reciprocity condition.

Study 2 introduced the SoVT and the results show functional neural and affective changes related to compassion training. The SoVT is composed of three video sets showing others in distress or everyday life situations. These video sets were matched based on a behavioral study with 265 healthy volunteers. A subsequent study with 94 female participants confirmed the matching and validated the SoVT on the neural and behavioral level. Using the SoVT, neural plasticity related to compassion training (*n* = 28) was studied in a pre-post intervention study with a matched active control group (*n* = 30). On the level of self-reported affect, compassion training increased empathy and positive affect, the latter even in response to others in distress. On the neural level, evidence for neural plasticity induced by compassion training was found in the medial orbitofrontal cortex, the putamen, the pallidum and the ventral tegmental area/ substantia nigra (VTA/SN). Converging evidence for the reliability of this network stemmed from two further independent compassion training studies (total *N* = 46). In addition, activations in these regions were observed in an expert meditation practitioner immersed in compassionate states.

To dissect the neural substrates of empathy and compassion in **Study 3**, one group of participants was first trained in empathy and subsequently in compassion (*n* = 25). The comparison with an active control group (*n* = 28) provided evidence for neural plasticity in empathy and showed a clear dissociation between empathy and compassion on the neural and affective level. Empathy increased negative affect and neural activity in the insula and the anterior midcingulate cortex, extending cross-sectional meta-analytic findings that have shown the involvement of these regions in empathy for pain. Subsequent training of compassion reversed the increase in negative affect and fostered positive affect along with non-overlapping changes in a neural network that spans the medial orbitofrontal cortex, the ventral striatum and the pregenual anterior cingulate cortex. These findings extend the results from Study 2, as well as previous cross-sectional studies of compassion, by showing that, even after an augmentation in empathic distress, compassion training can induce a shift towards positive emotions.

In **Study 4**, the covariance of cortical thickness was mapped, seeding from the anterior insula and the anterior midcingulate cortex, using data from 94 participants of Study 2. Seed regions were based on a recent functional empathy for pain meta-analysis. Firstly, functional connectivity patterns previously observed for these regions were be replicated. Secondly, structural covariance was related to trait and state measures of empathy. The results revealed that empathic states (empathy ratings in response to the SoVT) parametrically modulated anterior insula connectivity with inferior-frontal regions. This finding was specific for empathy as no such relation was observed for negative affective states in response to the SoVT. A similar, albeit weaker, relation was observed for questionnaire-derived empathic traits. This finding suggests that structural covariance analysis is a sensitive tool for mapping neural networks underlying social emotions.

In summary, the results obtained point to plasticity in the social brain and to the clear distinction of empathy and compassion. Compassion emerged as a new potent coping strategy that benefits the self through strengthening resilience and helps others through

increasing prosocial behavior. Finally, the presented data show that structural covariance analysis can determine specific neural networks related to empathy.

Zusammenfassung Deutsch

Das hauptsächliche Ziel dieser Dissertation war, Plastizität im sozialen Gehirn zu untersuchen und zu testen, welche Veränderungen in Verhalten, Affekt und neuronaler Aktivierung durch das Trainieren von Mitgefühl und Empathie bewirkt werden können. Um längsschnittliche Veränderungen in diesem Bereich reliabel zu messen, wurden zwei neue Aufgaben entwickelt und validiert: das Zurich Prosocial Game (ZPG) als neues implizites Maß für Hilfeverhalten und der Socio-affective Video Task (SoVT) als Maß für neuronale und affektive Reaktionen auf soziale Stimuli.

In **Studie 1** wurde das ZPG an 68 weiblichen Probanden validiert. Die Ergebnisse zeigen, dass das Spiel sensitiv ist für die Manipulation von Reziprozität, Hilfekosten und Stresssignalen. Unter Verwendung des ZPG in einer kurzen Interventionsstudie hat sich gezeigt, dass das Trainieren von Mitgefühl (n = 27), nicht aber das Trainieren von Gedächtnis (n = 32) zu einem Anstieg in Hilfeverhalten in einer Aufgabe führt, die keinen Bezug zum Training hat. Darüber hinaus korrelierten in der Mitgefühlsgruppe interindividuelle Unterschiede in der Übungsdauer positiv mit Veränderungen in nicht reziprokem Hilfeverhalten.

Studie 2 stellt den SoVT vor und zeigt funktionelle neuronale und affektive Veränderungen, die mit dem Trainieren von Mitgefühl zusammenhängen. Der SoVT besteht aus drei Video Sets, die das Leid anderer Menschen zeigen oder andere Menschen in Alltagssituationen darstellen. Diese Video Sets wurden anhand einer behavioralen Studie mit 265 Probanden gematcht. Eine darauf folgende Studie mit 94 weiblichen Probanden bestätigte die Parallelität der drei Video Sets und validierte den SoVT auf dem Level von neuronaler Aktivierung und Verhalten. Unter Verwendung des SoVT wurde neuronale Plastizität nach Mitgefühlstraining (n = 28) untersucht. Hierzu wurde eine Prä-Post Interventionsstudie mit einer aktiven Kontrollgruppe (n = 30) durchgeführt. Mitgefühlstraining führte zu einem Anstieg in selbst berichteter Empathie und positivem Affekt. Positiver Affekt stieg sogar in Reaktion auf das Leid anderer an. Auf neuronaler Ebene zeigte sich Evidenz für durch

Mitgefühlstraining induzierte neuronale Plastizität im medialen orbitofrontalen Cortex, Putamen, Pallidum und im ventralen Tegmentum/ der Substantia nigra. Konvergierende Evidenz für die Reliabilität dieses Netzwerks stammt aus zwei weiteren unabhängigen Mitgefühlsstudien (*N* = 46) die ebenfalls berichtet werden. Zudem wurde Aktivität in diesen Regionen auch in einem Meditationsexperten beobachtet, während dieser sich in Zustände des Mitgefühls vertiefte.

Um Unterschiede in den neuronalen Grundlagen von Empathie und Mitgefühl in **Studie 3** zu untersuchen, wurde eine Gruppe von Probanden zunächst in Empathie trainiert und anschließend in Mitgefühl (*n* = 25). Der Vergleich mit der aktiven Kontrollgruppe (*n* = 28) wies neuronale Plastizität von Empathie nach und zeigte, dass sich Empathie und Mitgefühl auf neuronaler und Verhaltensebene klar unterscheiden lassen. Empathie führte zu einem Anstieg in negativem Affekt und neuronaler Aktivierung in der Insula und im anterioren mittleren cingulären Cortex. Dies erweitert Befunde aus Quersschnittsstudien, in denen die zentrale Rolle dieser Regionen für Empathie mit Schmerz gezeigt wurde. Ein anschließendes Mitgefühlstraining kehrte den Anstieg in negativem Affekt um und erhöhte positiven Affekt. Zudem führte Mitgefühlstraining zu nicht überlappenden Veränderungen in einem neuronalen Netzwerk, das den mittleren orbitofrontalen Cortex, das ventrale Striatum und den vorderen Teil des anterioren cingulären Cortex umfasst. Diese Befunde erweitern sowohl die vorliegenden Resultate aus Studie 2, als auch Resultate aus früheren Querschnittsstudien von Mitgefühl, indem sie zeigen, dass Mitgefühlstraining sogar nach einem vorrausgehenden Anstieg in empathischem Stress eine Veränderung hin zu positiven Emotionen bewirken kann.

Studie 4 untersuchte bei 94 Probanden aus Studie 2 die Kovarianz der kortikal Dicke unter Verwendung von Ursprungsregionen in der anterioren Insel und dem anterioren mittleren cinuglären Cortex. Die Ursprungsregionen wurden an Hand einer früheren Meta-analyse von funktionellen Daten zu Empathie für Schmerz definiert. In einem ersten Schritt konnten funktionelle Konnektivitätsbefunde dieser zwei Regionen repliziert werden. In einem zweiten

Schritt wurde getestet, inwiefern strukturelle Kovarianz durch Charakter- und Zustandsmaßen von Empathie moduliert wird. Die Ergebnisse zeigten, dass die Konnektivität der anterioren Insel mit inferior-frontalen Regionen parametrisch durch empathische Zustände (Empathieeinschätzungen als Reaktion auf den SoVT) moduliert wird. Dieser Befund war spezifisch für Empathie, da für negativen Affekt als Reaktion auf den SoVT kein solcher Zusammenhang beobachtet wurde. Ein ähnlicher, allerdings schwächerer Zusammenhang wurde für die durch einen Fragebogen erhobene Charaktereigenschaft Empathie beobachtet. Dies deutet darauf hin, dass strukturelle Kovarianzanalyse ein sensitives Instrument darstellt mit dem abgebildet werden kann, welche neuronalen Netzwerke sozialen Emotionen zu Grunde liegen.

Zusammenfassend weisen die präsentierten Befunde auf Plastizität im sozialen Gehirn hin und zeigen einen klaren Unterschied zwischen Empathie und Mitgefühl. Mitgefühl zeichnete sich als neue und wirksame Copingstrategie ab. So hilft Mitgefühl einem selbst durch das Stärken von Resilienz und anderen durch das Fördern von prosozialem Verhalten. Zudem zeigen die Resultate, dass mit Hilfe von struktureller Kovarianzanalyse neuronale Netzwerke abgebildet werden können, die mit Empathie zusammenhängen.

Content

1. General Introduction

Outline of the Thesis

The principal aim of the present thesis was to study the neural, behavioral, and subjective plasticity of the social emotions empathy and compassion. The studies performed and the respective theoretical background, methods, results, and discussion are presented in a cumulative format, with each study corresponding to one publication. To place the present research in a broader context, this thesis begins with a General Introduction, which defines empathy and compassion, and related concepts. It also summarizes key findings in empathy research from the domain of psychology and more recent insights into the neural substrates of empathy from social neuroscience. This General Introduction is part of another publication written in the context of this thesis: a chapter for the Handbook of Human Affective Neuroscience.

Despite the impressive advances in empathy research over the last few years, the neural substrates of empathy are still poorly dissociated from other related emotions, like negative affect in general and compassion in particular. Moreover, the question of plasticity in social emotions has been neglected for a long time. One exception is the cross-sectional study in the domain of compassion expertise, in which expert meditation practitioners and novice meditation practitioners' compassionate responses to distressing sounds were compared (Lutz, Brefczynski-Lewis, Johnstone, & Davidson, 2008). In contrast, cross-sectional (comparing experts to non-experts) and longitudinal research of neural plasticity in other domains such as cognitive (Draganski et al., 2006; Olesen, Westerberg, & Klingberg, 2004), musical (Taubert, Lohmann, Margulies, Villringer, & Ragert, 2011), language (Guimaraes et al., 1998) and motor skills (Draganski et al., 2004; Karni et al., 1995) have received much greater attention over the last few decades.

To study affective plasticity of social emotions such as empathy and compassion, we first developed and validated two new tasks. One task was designed to repeatedly measure prosocial behavior in a laboratory setting, the other task was developed to repeatedly measure affective experiences and brain responses during confrontation with the distress of others. After developing and validating these tasks, we used them to study training-induced behavioral and

functional neural plasticity, as well as structural neural substrates of compassion and empathy. To this end, we employed functional magnetic resonance imaging (fMRI) and structural covariance analysis. These two neuroscientific methods are described in the last part of the introduction, titled “Methods: Magnetic Resonance Imaging”.

The development and validation of our novel, ecologically valid measure of prosocialness, the Zurich Prosocial Game (ZPG), is described in Study 1. The ZPG was developed to dissociate the influence of reciprocity, distress, and helping cost on prosocial behavior. In a longitudinal pre-post study, we used the ZPG to test whether training compassion, a social emotion related to prosocial motivation by definition (e.g., Keltner & Goetz, 2007), can increase prosocial behavior.

Study 2 introduced the Socio-affective Video Task (SoVT), which was constructed to reliably measure social affect and related neural function in longitudinal designs. Because retest effects like habituation often confound repeated measurements, the SoVT was composed of three parallel video sets (A, B, and C) to be used interchangeably. Extensive validations on the affective and neural level were conducted to assure the task’s reliability and validity. Using the SoVT, we investigated the degree to which training compassion induces functional neural plasticity, and changes in positive affect, negative affect, and empathy. We expected that compassion-training, but not active memory control training, would augment neural activity in regions previously observed in cross-sectional studies of compassion, love, and affiliation, such as the medial orbitofrontal cortex, the striatum, and the ventral tegmental area/substantia nigra (VTA/SN; e.g., Bartels & Zeki, 2004; Beauregard, Courtemanche, Paquette, & St-Pierre, 2009; J. W. Kim et al., 2009; Strathearn, Fonagy, Amico, & Montague, 2009). Furthermore, we hypothesized that compassion training would decrease negative affect and increase positive affect.

Another objective was to study functional neural plasticity related to empathy training and to delineate it from plasticity related to compassion training (Study 3). To achieve this goal, a group of participants first of all completed baseline pre-training measurements. They were then trained in empathy and subjected to post-training measurements. After that, they received compassion training, which was followed by a second post-training measurement. We expected that, compared to the pre-training measurement, empathy training would augment negative affect and empathy, and induce functional neural plasticity in core structures underlying empathy for pain, such as the anterior insula (AI) and the anterior medial cingulate cortex,

(aMCC; Fan & Han, 2008; Lamm, Decety, & Singer, 2011). Conversely, compassion training was expected to reverse these effects, induce positive emotions, and augment activations related to compassion.

Finally, we investigated the structural covariance of the empathy-related regions: the AI and the aMCC (Study 4). Functional neural signatures of self-experienced pain and empathy for pain have been shown to overlap in regions like the AI, both on the level of general linear models (Lamm et al., 2011) and multi-voxel pattern classification (Corradi-Dell'Acqua, Hofstetter, & Vuilleumier, 2011). In the spirit of previous functional analyses, which have shown distinct connectivity patterns of the AI and the aMCC for self-experienced pain and empathy for pain (Zaki, Ochsner, Hanelin, Wager, & Mackey, 2007), we used covariance analysis to disentangle the structural connectivity patterns specific to empathic traits and states. Furthermore, we dissociated the latter from negative affective states. The findings obtained from all the studies conducted are summarized, integrated and critically reflected upon in the General Discussion, which ends with an outlook for future research directions.

Empathy from the Perspective of Social Neuroscience

Corresponding publication:

Klimecki, O., & Singer, T. (2013). Empathy from the perspective of social neuroscience. In: J. Armony & P. Vuilleumier (Eds.), *Handbook of Human Affective Neuroscience.* New York: Cambridge University Press. With kind permission of Cambridge University Press.

Abstract

One of the main aims in the field of social neuroscience is to unravel the neural underpinnings of empathy, which can be defined as the ability to share the emotional experience of others. We begin by defining empathy and delineating it from related concepts such as theory of mind and action understanding on the one hand and compassion and empathic distress on the other hand. Next, we review psychological findings on the relation between empathy and prosociality and move to the question of *how* we are able to empathize with other people. Findings from social neuroscience on empathy for pain and other modalities suggest that we use shared neural networks to simulate the emotional experience of others. We discuss how neural networks for self-experienced pain and empathy for pain are shared but distinct and specifically emphasize the role of the insula in this context. We conclude the chapter by summarizing modulating factors of empathy and outlining future research goals in the field.

Empathy, which can be broadly defined as the capacity to share and understand other people's emotions (for comprehensive reviews, see Batson, 2009a; de Vignemont & Singer, 2006; Decety & Jackson, 2006; Eisenberg, 2000; Hoffman, 2000; Singer & Lamm, 2009; Singer & Leiberg, 2009), has recently become an important focus of attention in the field of social neurosciences. What motivates the quest for the neural substrates underlying our understanding of emotions in others?

After many years in which neuroscientific research mainly focused on cognitive and sensory processing, attention has increasingly turned to understanding how the human brain tackles emotions as well as social interactions, which – after all – are both phenomena at the core of our existence as social beings. Thus, the field of social neuroscience has started to investigate the neural mechanisms underlying social cognition and emotions, such as our ability to empathize. In addition to the basic understanding of the biological mechanisms underlying social emotions and empathy in healthy individuals, research on the neural substrates of empathy may also help us understand clinical phenomena related to a lack in affective and social skills such as autism, which is characterized by impairments in social interaction and communication (American Psychiatric Association, 2000), or alexithymia, a subclinical phenomenon associated with difficulties in identifying and describing emotions (Nemiah, Freyberger, & Sifneos, 1976).

The importance of empathy in our everyday lives becomes clear when we try to imagine what it would be like to live in a world completely devoid of empathy. Take the following scenario, for instance: a baby starts crying while her mother is reading a book. Without empathy, the mother would probably continue reading and not look after the baby. Her capacity to empathize, however, enables her to realize the baby's needs and react to them appropriately. As this example illustrates, empathy does not only motivate other-related prosocial behavior but also enables us to better predict the behavior of others and adapt our reactions accordingly. Finally, empathy also plays a crucial role in observational learning – by witnessing the emotional reaction of others in different circumstances, we learn which situations are good for us and which situations are better avoided.

In this chapter, we start out by revisiting the definition of empathy and delineating it from other routes to social understanding, namely, theory of mind and action understanding. We will then move on to examine the theoretical and neural underpinnings of concepts like emotion contagion and mimicry, which can be thought of as antecedents of empathy, and compassion and empathic distress, which are introduced as consequences of empathy. Before turning our focus to how research in social neuroscience has advanced our understanding of empathy in

the human brain, we review the major contributions of psychological research to our understanding of empathy and its relation to prosocial behavior. Since the neural underpinnings of empathy have most prominently been examined in the domain of empathy for pain, we will start out by summarizing this line of research and discussing the reported results in light of the shared network hypothesis. In this context, we will stress the specific role of the insula as a neural structure that processes both interoception and empathy. Subsequently, we will describe factors that modulate the experience of empathy for pain along with their neural underpinnings, before turning to neural correlates of empathy in other domains such as touch or smell. Finally, we will show initial findings from social neuroscience research focusing on more positive aspects of empathy, such as compassion. We conclude the chapter by outlining outstanding questions in the field.

Defining Empathy and Related Concepts

Empathy is commonly defined as the human capacity to understand and share another person's emotion without confusing it with one's own emotional state (for comprehensive reviews, see Batson, 2009a; de Vignemont & Singer, 2006; Decety & Jackson, 2006; Eisenberg, 2000; Hoffman, 2000; Singer & Lamm, 2009; Singer & Leiberg, 2009). In other words, we empathize with another human being when we vicariously share their affective state but at the same time are aware that it is the other person's emotion which is causing our response.

In this section, we will first point out the conceptual difference between *empathy*, *mentalizing* and *action understanding*, which can be conceived as different routes to the understanding of others. After showing that the psychological distinction between these three concepts is paralleled by differences in the underlying neural networks, we will have a closer look at the "sisters of empathy", namely *emotion contagion*, *mimicry, sympathy and compassion*, all of which are concepts closely related to empathy.

Mentalizing and Action Understanding as Alternative Routes for Understanding Others

Besides *empathy*, which can be seen as the emotional route for understanding others, there are at least two other ways of putting oneself into another person's shoes. On the one hand, we have the cognitive ability to understand the thoughts, beliefs and intentions of others, which is called *mentalizing, perspective taking* or *theory of mind* (*ToM*) (Frith and Frith, 2003; Premack and Woodruff, 1978). On the other hand, we have the capacity to understand the motor intentions of others, which has been associated with the discovery of mirror neurons (see Rizzolatti & Sinigaglia, 2010, for a review). Although all three often occur simultaneously in

everyday social cognition, the psychological and neural processes underlying these distinct routes to understanding others can be clearly distinguished (for reviews, see de Vignemont & Singer, 2006; Preston & de Waal, 2002; Singer & Lamm, 2009). Cognitive processes related to *theory of mind* have been associated with activations in the medial prefrontal cortex (mPFC), superior temporal sulcus (STS) and the adjacent temporoparietal junction (TPJ) (for reviews, see Amodio & Frith, 2006; Frith and Frith, 2006; Saxe, 2006; Saxe & Baron-Cohen, 2006; Mitchell, 2009;), whereas the neural correlates of *action understanding* are found in a neural network spanning the inferior parietal lobe (IPL), the inferior frontal gyrus and ventral premotor areas (see Rizzolatti & Sinigaglia, 2010, for a recent review). In monkeys, recordings from so-called *mirror neurons* in corresponding areas have revealed that these neurons encode both, the execution of an action and the observation of the same action in others (Gallese et al., 1996; Rizzolatti et al., 1996). Paralleling the establishment of this mirror network in monkeys, recent research has extended these findings to humans by means of magnetoencephalography, (MEG; Hari et al., 1998), transcranial magnetic stimulation (TMS; Cattaneo et al., 2010; Fadiga et al., 1995) and functional magnetic resonance imaging (fMRI) studies (Iacoboni et al., 2005), suggesting that monkeys as well as humans may use the same neural structures to encode their own actions and to understand the actions of others (for a review, see Rizzolatti & Craighero, 2004, or Rizzolatti & Sinigaglia, 2010). In a meta-analysis, Grèzes and Decety (2001) compared the activation foci of a variety of studies and found that overlapping activations for execution, simulation and observation of actions are located in the supplementary motor area, dorsal premotor cortex, supramarginal gyrus and superior parietal lobe. Finally, as discussed in detail below, the neural correlates of empathy are mainly observed in limbic and paralimbic areas such as the anterior insula (AI) and anterior cingulate cortex (ACC) (for recent reviews, see Lamm & Singer, 2010; Singer & Lamm, 2009; Singer & Leiberg, 2009). Together, theory of mind, action understanding and empathy allow us to infer the thoughts, motor intentions and emotions of others, thereby facilitating social interactions.

Different Components of Empathy (Figure 1.1)

After having introduced theory of mind, action understanding and empathy as three complementary routes to the understanding of others relying on distinct neural networks, we will now describe the different facets of empathy-related phenomena. They range from rather automatic and primitive reactions, such as mimicry and emotional contagion (which can be thought of as precedents of empathy), to states like compassion or empathic distress which

follow from empathy and are, themselves, important determinants of behavior (Batson, 2009b; de Vignemont & Singer, 2006; Eisenberg, 2000; Goetz et al., 2010; Klimecki & Singer, 2012; Singer & Lamm, 2009).

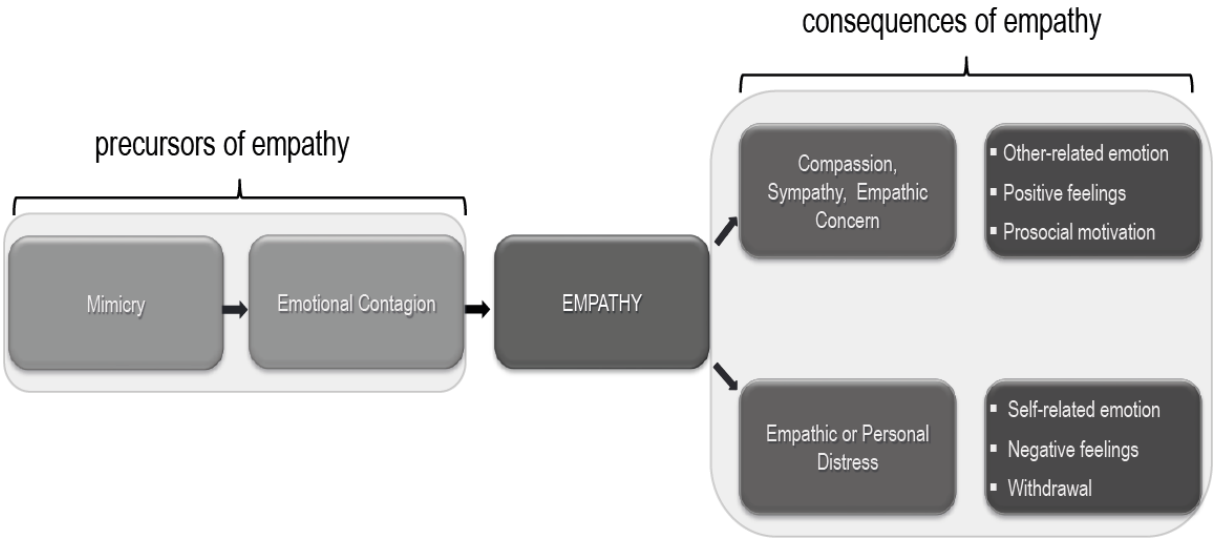


Figure 1.1. Schematic model showing the precursors and consequences of empathy.

Mimicry and Emotional Contagion

Mimicry can be described as an automatically elicited response mirroring another person's emotional expression conveyed by facial, vocal or postural expressions or by movements (see Hatfield et al., 2009, for a recent review). In the domain of facial mimicry, for instance, electromyographic (EMG) recordings reveal that the visual presentation of emotional faces elicits corresponding emotional facial expressions in the observer (see Dimberg & Öhman, 1996, for a review); the perception of happy faces evokes increased activity in the *zygomatic major* muscle (which raises the corners of the mouth during smiling), whereas the perception of angry faces leads to increased activity in the *corrugator supercilii* muscle (which is associated with frowning). The short latency between stimulus onset and facial reaction (300-400 ms) supports the claim that facial mimicry is elicited automatically and occurs preattentively. Complementing these findings in the domain of facial mimicry, researchers observed that people also tend to synchronize their vocal expressions and adopt the postures and movements of others (reviewed for example by Hatfield et al., 2009).

Emotional contagion goes one step further than mimicry in that the automatic imitation and synchronization of displayed emotions – be it at the level of facial expressions, vocalizations, postures or movements – results in a convergence in the actual emotional experience (see Hatfield et al., 2009, for recent review, or Dimberg & Öhman, 1996). This is in line with the claim that peripheral physiological feedback shapes our emotional experience (e.g., Adelmann & Zajonc, 1989) and the finding that perceiving another person in a certain emotional state can induce a congruent state in the observer (e.g., Gottman & Levenson, 1985; Harrison et al., 2006; Neumann & Strack, 2000). More concretely, this suggests that seeing another person smile makes us smile, which in turn makes us feel happy. Mimicry and emotional contagion can thus be regarded as precursors of empathy. In contrast to emotional contagion and mimicry, however, which can occur without self-other distinction, empathy crucially relies on the capacity to distinguish between oneself and the other. In other words, an empathic observer is aware that he or she is experiencing feelings vicariously and that these feelings were induced by emotions experienced by another person, not by his or her primary experience (de Vignemont & Singer, 2006).

Empathic Distress

Whereas mimicry and emotion contagion can be regarded as precursors of empathy, empathy may, in turn, lead to two opposing consequences, namely empathic distress and compassion. *Empathic distress*, which is also referred to as *personal distress*, is an aversive and self-oriented emotional response to the suffering of others. It often results in withdrawal behavior, which is motivated by the desire to protect oneself from negative emotions (Batson et al., 1983; Eisenberg et al., 1989). In this light, empathic distress, although regarded as a consequence of empathy, falls somewhat between emotion contagion and empathy as the self-other distinction becomes blurred when the secondary empathic experience triggered by another person's suffering becomes so overwhelming that it turns into distress. This concept is particularly important for people working in health care sectors, where repeated encounters with suffering may lead to burnout (for review, see Klimecki & Singer, 2012). Therefore, it is vital to find alternative ways of dealing with the suffering of others. Essentially, empathy does not necessarily have to take the route of empathic distress, but can instead lead to more adaptive positive emotions of compassion.

Compassion, Empathic Concern and Sympathy

Compassion, empathic concern and sympathy[1] all denote affective states that can be experienced as a consequence of empathy and that are not shared *with* someone, but felt *for* someone (Batson, 2009a; Singer & Lamm, 2009). Empathy or "*feeling with*" denotes a state in which the feelings of someone else are vicariously shared so that the empathizer feels an isomorphic state (de Vignemont & Singer, 2006): upon witnessing a sad person, the empathizer becomes sad himself. On the other hand, "*feeling for*" someone refers to an emotional state that is not necessarily isomorphic to the target's affective state, but instead relies on feelings of concern for the other. More precisely, compassion can be defined as "*the emotion one experiences when feeling concern for another's suffering and desiring to enhance that individual's welfare*" (Keltner & Goetz, 2007). In other words, compassion consists of two main components. First, there is a caring feeling for the suffering of others which secondly motivates behavior that is aimed at relieving the other's suffering. As will be described in detail below, research in social and developmental psychology has demonstrated that empathic concern actually motivates prosociality and helping behaviors (see Batson, 2009b, or Eisenberg, 2000, for a review). In sum, whereas empathy is a vicarious emotion isomorphic to the emotional experience of the other, the experience of compassion, sympathy or empathic concern denotes an affective state experienced with regard to the other that does not encompass the sharing of negative affect but instead relies on a feeling of care and concern that motivates prosocial behavior. Before turning to empirical evidence for the promotion of prosocial behavior by compassion, we will provide an overview of measures developed by psychologists to assess empathy and its different components.

Empathy Research in Psychology

Empathy and its relation to prosocial behavior have been investigated by psychologists from various fields. In the following passage, we will first describe self-report, behavioral and physiological measures of empathy before summarizing how research in developmental and social psychology established the link between empathy and prosocial behavior. Theodor Lipps (1903) first introduced the concept of empathy by proposing that we imitate the gestures and actions of others in order to understand their inner states. In order to measure empathy, Davis

[1] Since these three terms are often used to denote the same underlying concept (Batson, 2009a) - with Batson primarily using the term empathic concern (Batson, 2009b) while Eisenberg rather speaks of sympathy (Eisenberg, 2000) - we will use these terms interchangeably throughout the chapter.

(1980) developed a questionnaire called Interpersonal Reactivity Index (IRI) that includes four distinct components: perspective taking, empathic concern, personal distress and fantasy. Perspective taking is very close to the above-mentioned notion of theory of mind as it measures the tendency of people to cognitively adopt the perspective of others. Empathic concern, however, is more closely related to the concept of sympathy or compassion as discussed above, while the personal distress subscale measures how prone individuals are to experiencing discomfort as a result of witnessing distress in others. Finally, the fantasy scale asks people how well they tend to identify with fictional characters in books or movies. The Balanced Emotional Empathy Scale (BEES; Mehrabian, 1997; Mehrabian & Epstein, 1972), which includes subscales such as "susceptibility to emotional contagion" or "sympathetic tendency", is another questionnaire that measures the emotional aspects of empathy.

These self-report measures are complemented by the assessment of empathic accuracy. Levenson and Ruef (1992), for instance, presented subjects with videotaped marital interactions. By comparing the subject's ratings of positive and negative emotions displayed in the video with the self-report of the actual target, an empathy accuracy score is obtained that assesses the degree to which subjects correctly identify the emotional state of others. Additionally, physiological measures such as skin conductance and heart rate can be compared with respect to the level of physiological linkage between the target and the subject empathizing with the target (see also Gottman & Levenson, 1985). Whereas these measures focus more on the emotional and physiological linkage, Ickes and colleagues (see Ickes, 1993, for a review) used a similar approach to determine to what degree subjects succeed in inferring another person's thoughts and the content of their emotions.

Empathy and its Relation to Prosocial Behavior

In order to establish the link between empathy and prosocial behavior, researchers in developmental and social psychology conducted several studies, which suggest that empathy can have two opposing consequences depending on the nature of the empathic experience. In the field of social psychology, Batson and colleagues conducted several experiments (see Batson et al. 1981, 1983, 1987), which established that participants experiencing empathic concern feel the urge to help people in need, irrespective of whether the adverse situation is easy or difficult to escape. On the contrary, participants suffering from empathic distress, tend to be more self-oriented and to withdraw from negative experiences whenever possible and thus only choose to help when the aversive situation is difficult to escape. This tendency to escape

might result from a self-oriented aim to protect oneself by reducing one's own negative affect. The link between empathic concern vs. distress and prosocial behavior has been extended to children by research in developmental psychology, which in addition, established a relationship between the physiological correlates of emotional reactions and helping behavior (for a review, see Eisenberg, 2000). Eisenberg and colleagues (1989), for instance, found that adults' self-reports and facial display of sympathy predicted prosocial behavior. In children, however, the propensity to offer help was not predicted by their verbal reports of distress and sympathy. Instead, children's facial display of distress was negatively related to helping. This difference between adults and children may indicate that children's capacity to report their emotional experience reliably is underdeveloped. In sum, these results indicate that although there are differences between adults and children, empathic concern (or sympathy) promotes prosociality whereas empathic distress is associated with withdrawal tendencies. Interestingly, a recent study revealed that prosocial behavior towards strangers can be increased by short-term training of compassion (Leiberg et al., 2011). These findings have broad implications for the implementation of compassion training in schools and other public organizations, since compassion was shown to be a trainable and generalizable skill, motivating prosocial behavior that even extends towards strangers.

Empathy in Social Neuroscience

After having reviewed the various methods employed by social and developmental psychologists to study empathy and related concepts, we turn to the paradigms developed by social neuroscientists. Since a vast majority of empathy studies in the fMRI setting are based on variations of the empathy-for-pain paradigm, we will begin this section by describing the main methods, findings and implications of empathy-for-pain research. Subsequently, we will integrate these findings with neuroscientific studies on empathy in various other domains.

Empathy for Pain

Given that empathy is a highly social phenomenon, researchers were (and still are) faced with the difficulty of coming up with a paradigm that is compatible with fMRI measurements while at the same time being ecologically valid. To reconcile both aims, Singer and colleagues (Singer et al., 2004) designed an empathy for pain paradigm in which two participants present in the same scanner environment alternately receive painful stimuli administered through electrodes attached to the back of their hand. More specifically, both the person lying in the scanner and the person sitting next to the scanner receive painful and non-painful stimulation. This setup

allows for the comparison between brain responses elicited when the scanned subjects experience painful stimulation with neural activations related to witnessing another person experiencing pain. During the brain scan, arrows in different colors indicate to both participants which of them is going to be stimulated next and whether the stimulation is going to be painful or not. Under the assumption that people in a very close relation feel strong empathy for each other, Singer and colleagues (2004) started out by examining empathy for pain in couples. Intriguingly, the results of this study show that the neural signature of empathy for pain is very similar to the neural processes underlying the self-experience of pain (for brain circuits mediating pain perception, the so-called 'pain matrix'). More specifically, empathy for pain activated selective parts of the neural pain matrix including AI and ACC, which both are key regions in processing bodily and feeling states (see also Singer et al., 2009). These findings have been replicated by several studies using similar paradigms (Bird et al., 2010; Hein et al., 2010; Singer et al., 2006, 2008). Converging evidence for the involvement of AI and, less consistently, ACC in empathy for pain also has been obtained from several other studies (some of which will be discussed in more detail below) using various paradigms ranging from simultaneous pain administration in the scanner to the presentation of photographs or videos depicting painful events (e.g., Botvinick et al., 2005; Cheng et al., 2007, 2010; Danziger et al., 2009; Decety et al., 2010; Gu & Han, 2007; Jackson et al., 2006; Lamm et al., 2007a, 2007b; Lamm & Decety, 2008; Moriguchi et al., 2007; Morrison, Lloyd, di Pellegrino, & Roberts, 2004; Ogino et al., 2007; Saarela et al., 2007; Zaki et al., 2007).

In support of the claim that the observed neural activation patterns are closely related to the concept of empathy developed by psychologists, Singer and colleagues (2004) showed that higher self-reports of empathy as measured by the BEES and the IRI were accompanied by increased neural activity during empathy for pain in the left AI and ACC, speaking for the external validity of the empathy-for-pain paradigm. Further evidence for the link between the observed brain responses and empathic experience stems from studies in which neural activation patterns in specific empathy-related regions were shown to correlate with self-reported impressions and actual helping behavior (Hein et al., 2010), as well as with individual unpleasantness ratings (e.g., Jabbi et al., 2007; Lamm et al., 2007b; Saarela et al., 2007; Singer et al., 2008). Importantly, the existence of a shared neural network for self-experienced pain and empathy for pain is supported by a recent meta-analysis of Lamm and colleagues (2011), which shows consistent overlaps in AI and ACC across nine independent fMRI studies and will be discussed in more detail below. Although the above mentioned results point to a shared neural representation of self-experienced and vicariously experienced pain, the question still remains

whether activations in the AI and the ACC are overlapping on the level of neuronal subpopulations and single neurons (e.g., Singer & Lamm, 2009).

Shared and Distinct Neural Networks in Empathy for Pain

In addition to confirming an overlap between first-hand and vicarious pain experience in AI and ACC, the meta-analysis of Lamm and others (2011) suggests that this shared network can be accessed via several routes depending on the paradigm employed. Whereas the use of picture-based empathy paradigms is linked to additional activation increases in the inferior parietal, ventral premotor and dorsomedial cortex (a neural circuitry typically observed in action understanding), cue-based paradigms induce activation in networks typically linked to theory of mind, such as mPFC, precuneus, STS and TPJ.

Distinct connectivity patterns of the AI and ACC with other brain regions during self-experienced pain versus empathy for pain have been examined by Zaki and colleagues (2007), who reported that self-experienced pain is associated with stronger connectivity between the AI and regions involved in the transmission of painful sensations such as clusters in the midbrain, periaqueductal gray and mid insula. Conversely, empathy for pain has been shown to be associated with higher connectivity of ACC and AI with brain regions implicated in social cognition and affect processing, such as the medial prefrontal cortex.

Converging evidence for the activation of sensory brain structures during self-pain stems from the above mentioned meta-analysis (Lamm et al., 2011), where the direct experience of pain has been shown to recruit mid- and posterior insula, as well as primary sensory cortices in addition to AI and ACC (see Figure 1.2 for illustration). The stronger involvement of brain regions processing sensory information in self-experienced pain suggests that we share the pain of others by accessing the neural structures representing our own affective states, while leaving aside sensory and nociceptive components.

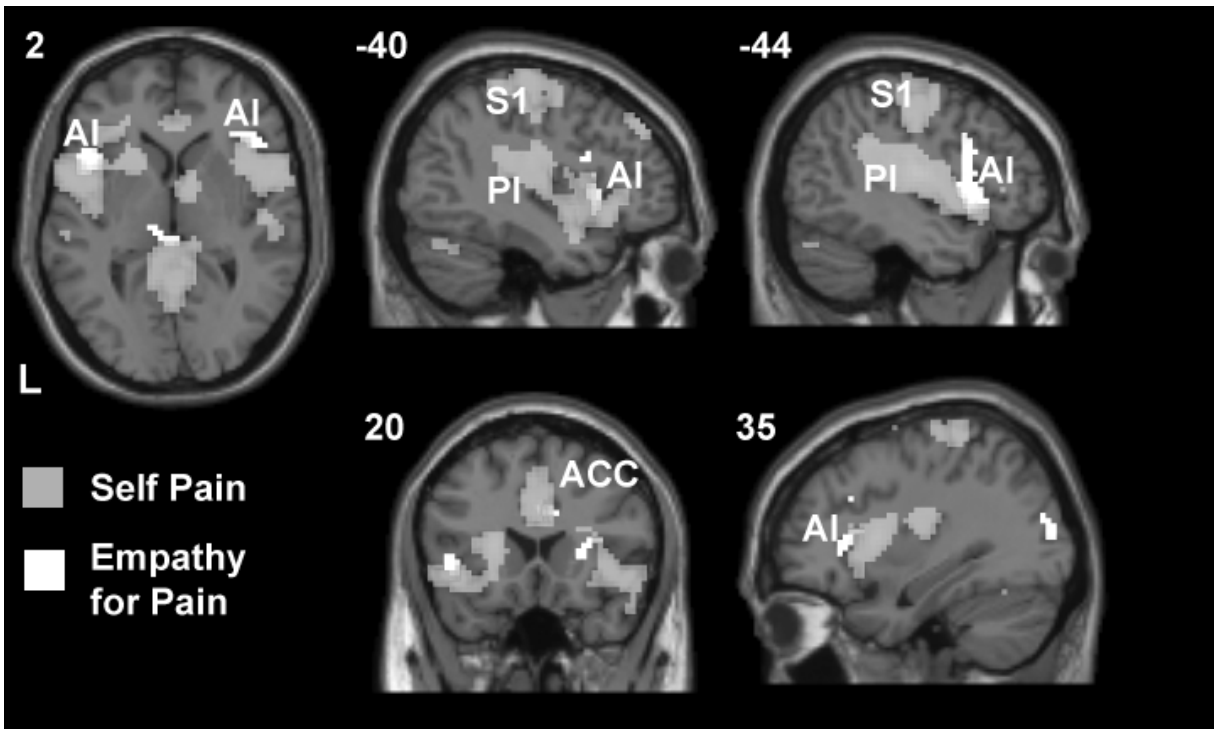


Figure 1.2. Shared and distinct neural networks for self experienced pain and empathy for pain. Depicted functional neural activations are the result of a meta-analysis based on nine fMRI studies investigating empathy for pain (Lamm et al., 2011). Activations related to self-experienced pain (green) encompass a large portion of the insula, including the middle and posterior insular cortex, whereas activations related to empathy for pain (red) are restricted to the most anterior parts of AI, where they overlap with activations related to self-experienced pain. Functional activation maps are overlaid on a high-resolution structural MRI scan in standard stereotactic space (MNI space). White labels indicate slice number in stereotactic space, L = left hemisphere, AI = anterior insula, ACC = anterior cingulated cortex, PI = posterior insula, S1 = primary somatosensory cortex. With kind permission from Springer Science and Business Media: Springer, Brain Structure and Function, 214, 2010, p.581, The role of anterior insular cortex in social emotions, Lamm, C. & Singer, T., Figure 1.

Moreover, the meta-analysis on empathy for pain showed that contralateral SI activation, which most likely encoded somatosensory aspects of painful sensations, is restricted to self-experienced pain in cue-based studies. Picture-based studies, on the contrary, evoke bilateral SI activations during both, empathy for painful and non-painful situations, thus pointing to an unspecific role of the primary somatosensory cortex in empathy for pain which is presumably

related to seeing body parts being touched. Interestingly, activity in somatosensory cortices can also be increased when subjects are instructed to evaluate the sensory consequences of painful stimuli, suggesting that attention can influence the quality of the empathic experience and that this shift in focus is accompanied by corresponding brain activations (Lamm et al., 2007b). Complementing these findings, sensorimotor resonance with empathy for visually depicted pain has been shown by studies using TMS (Avenanti et al., 2005, 2006) and EEG (Bufalari et al., 2007; Valeriani et al., 2008). Taken together, these results point to a core shared network for self-experienced pain and empathy for pain in the ACC and AI that varies in its connectivity with other brain regions. Thus, AI and ACC are co-activated with areas involved in processing self-related components of nociceptive experience when experiencing pain in one's self. However, when empathizing with others, AI and ACC are co-activated with networks involved in social cognition (ToM and action observation). This suggests that the information available in the task and the situational demands determines which of the social cognition networks will be predominantly engaged.

With regard to the quality of the empathic experience, the reported studies suggest that we primarily share painful experiences by simulating the affective and not so much the sensory and nociceptive components of pain. Note, however, that there seems to be a graded activation of the neural pain matrix, such that activations in posterior insula and secondary somatosensory cortex can be observed although no pain is administered if subjects adopt a first-person perspective (Jackson, 2006) or simply imagine a painful event (Ogino et al., 2007). Since these activation patterns do not include the primary somatosensory cortex, there seems to be a continuum between directly and vicariously experienced pain that is reflected in the underlying neural substrates.

The Role of Insula in Empathy and Interoception

The posterior-anterior gradient of the insula in self-experienced pain and empathy for pain is mirrored by research showing that primary nociceptive information is first processed in dorsal posterior parts of the insula before being remapped and integrated with other information in the AI where emotions access consciousness (Craig 2002, 2009). These results fit into the general notion of the insula as a key player in processing interoceptive information from the body (Craig, 2002; Damasio, 1994; Ostrowsky et al., 2000, 2002) and emotions more generally (shown by a meta-analysis by Kober et al., 2008). Given these findings, Singer et al. (2004, 2009) proposed that the insula fulfills a dual role: a) processing bodily information, such as heart beat or temperature-related sensations, which are then integrated into global feeling states and b)

predicting the affective states of others in the process of empathizing. In other words, we use our own bodily and affective representation to understand the emotional experiences of others. This interpretation implies that a deficit in understanding our own feelings should entail difficulties in empathizing with the feelings of others. Indeed, this claim was confirmed in studies focusing on people with alexithymia, a subclinical phenomenon characterized by difficulties in identifying and describing emotions (Nemiah, Freyberger, & Sifneos, 1976). Silani and colleagues (2008) demonstrated that activity in the AI during interoception on emotional stimuli diminished with increasing alexithymia scores, whereas activation in AI during interoception was positively related to trait empathy. Furthermore, this study showed that a higher degree of alexithymia was accompanied by lower levels of trait empathy. Extending these results, Bird and colleagues (2010) using the Singer et al. (2004) empathy for pain paradigm, found that AI activations are also decreased when highly alexithymic participants are asked to empathize with others in pain.

Modulation of Empathy

In the following section, we will discuss findings indicating that empathic brain responses in AI are not just modulated by person-specific characteristics such as alexithymia, but can also be affected by the relation to the target (e.g., liking or disliking), contextual and attentional factors, as well as by the appraisal of the situation (for recent reviews, see de Vignemont & Singer, 2006; Singer & Lamm, 2009). Several studies have shown that stimuli which under normal circumstances lead to empathic responses fail to induce empathy in certain situations and may even evoke the opposite, namely what has been referred to as “Schadenfreude” – the joy of witnessing another person’s misfortune.

Relationship between Empathizer and Target

Such a reversal of emotional response to the pain of others was, for example, observed by Singer and colleagues (2006) when they studied how perceived fairness affects empathy for pain. In their experiment, participants first played an economic game with two other volunteers, one of whom played fairly while the other played unfairly. When the participant was subsequently scanned, the fair and unfair players (who were actually confederates) were sitting next to the scanner and all three alternately received painful or non-painful stimulation on the backs of their hands. In line with previous findings, Singer and colleagues observed that both the first-hand experience of pain and empathy for pain in the case of the fair person rely on shared neural representations in the AI and the ACC. However, when it came to empathizing

with the pain of the unfair player, only female subjects showed greater activations in these regions. Male subjects, by contrast, showed a decline in AI activation while witnessing the unfair, as opposed to the fair, player in pain. This reduction in the men's neural empathy response was accompanied by increased activation in the nucleus accumbens – a region known to be crucially involved in reward processing (for recent reviews, see Knutson & Cooper, 2005; Schultz, 2000). Moreover, the extent of nucleus accumbens activation was positively correlated to the subjectively expressed desire of revenge. In other words, activity in this reward-related brain structure was higher when men reported a stronger desire for revenge towards the unfair player. This activation pattern may imply that men actually experienced Schadenfreude when witnessing the unfair player being punished.

In another study, Hein and colleagues (2010) extended these findings by showing a relationship between empathy-related brain responses in AI and subsequent prosocial behavior. The authors examined ingroup-outgroup biases in male soccer fans while these witnessed a fan of their favorite team (ingroup) or a fan of a rival team (outgroup) receiving painful electric shocks. As expected, the observation of ingroup members receiving pain was linked to greater AI activations. More importantly, the intensity of AI activation actually predicted the degree to which subjects would later help their ingroup member by taking the painful shocks themselves. In contrast, nucleus accumbens activation elicited by witnessing an outgroup member suffering predicted a refusal to help and reflected how negatively the subject evaluated the outgroup member. These findings imply that empathy-related insula activation drives altruistic behavior, whereas an antagonistic signal in nucleus accumbens reduces the propensity to help. Other examples of factors that influence the nature of social relationships, and thereby the degree of empathy and its neural correlates, are ethnicity (Xu et al., 2009) and closeness to the other person (Cheng et al., 2010).

Characteristics of the Empathizer

In a similar line as the above described alexithymia research, which established the link between high alexithymia (Silani et al., 2008) and low empathy, as well as the decrease of AI activations during empathy for pain in alexithymic participants (Bird et al., 2010), the characteristics of the empathizer have also been shown to influence empathic experiences in other domains. Cheng and colleagues (2007), for instance, showed that when observing needles being inserted into different body parts, participants without experience in acupuncture showed activations in the neural network involved in empathy for pain, whereas physicians who practice acupuncture themselves did not show such a neural response.

Situational Context, Attention and Appraisal

The role of contextual appraisal in relation to empathy has been examined in relation to the attribution of responsibility. Decety and colleagues (2010), for example, tested the degree to which empathic responses towards videos of pain in AIDS patients differed as a function of the target's responsibility (infection through transfusion or drug taking). The self-report measures and the neural activations of the subjects both conveyed that attributed responsibility influences the extent of empathic response. Participants reported higher pain and empathy ratings towards the pain of transfusion targets as compared to drug targets and these reports were accompanied by greater activation of the neural networks involved in processing pain (AI and ACC).

Another factor that has been shown to influence empathy is attention. Gu and Han (2007) report lower ACC activity when participants were asked to count neutral stimuli in images displaying painful events compared to rating the intensity of the pain. Finally, Lamm and colleagues (2007b) investigated the effects of cognitive appraisal and perspective taking on empathic responses by presenting video clips of painful facial expressions during a medical treatment. Perspective taking was varied by instructing the participants to imagine themselves in the depicted situation versus imagining the feelings of the patient during the medical treatment. In order to manipulate cognitive appraisal, subjects were either told that the medical treatment had been beneficial or unsuccessful. When subjects were told that the treatment had not been successful, they provided higher ratings of pain and unpleasantness than in the condition in which the treatment had been beneficial. Brain data corroborate these findings by showing stronger activations in the perigenual ACC for the ineffective treatment condition. With regard to the different perspectives adopted, results show that participants reported higher personal distress when they imagined themselves in the patient's situation (self-perspective), whereas participants reported more empathic concern when they cognitively differentiated between the patient and themselves (other-perspective). The observed increase of empathic distress when adopting a self-perspective fosters the claim that empathic distress arises due to an over-identification with the suffering of others. On the neural level, adopting a self-perspective was associated with increased activations of the neural pain matrix. This finding speaks for a stronger sharing of pain in distress and underscores the important distinction between empathic distress and empathic concern as two opposing outcomes of empathy associated with different qualities of emotional experience.

Empathy for Touch, Smell and Taste

As reviewed above, a majority of neuroscience studies on empathy have focused on pain. However, an early study focused on the examination of shared representations for smell and disgust (Wicker et al., 2003). In their experiment, Wicker and colleagues studied how processing the actual experience of disgusting olfactory stimuli differs from processing others' visual display of disgust. The results of this fMRI study show that both self- and other-related disgust are accompanied by overlapping activation in AI and ACC. Further support for this finding comes from a study by Jabbi and colleagues (2008), who showed AI activation irrespective of whether subjects tasted an unpleasant substance, viewed disgusted facial expressions or read disgusting scenarios. Finally, Keysers and others (2004) reported that, whereas the sensory experience of being touched is specifically linked to activation in the contralateral primary somatosensory cortex, the neural signatures of being touched and observing touch overlap in the secondary somatosensory cortex. In summary, these studies parallel findings on empathy in the domain of pain by providing evidence for the involvement of ACC and AI in empathy for other modalities such as touch, smell and taste.

The Compassionate Brain

Although social neuroscience has mostly focused on finding evidence for shared networks and their modulation so far, the field has recently moved forward to the investigation of positive consequences of empathy such as empathic concern, sympathy or compassion. Bartels and Zeki, for instance, studied the neural correlates of romantic love (2000) and maternal love (2004) by means of fMRI. Their results show that both types of love activate the middle insula, the dorsal part of the ACC and the striatum (comprised of the putamen, globus pallidus and caudate nucleus). Similar results were reported by Beauregard and colleagues (2009) who observed increased activations of the middle insula, the dorsal ACC, the globus pallidus and the caudate nucleus when their participants adopted a stance of unconditional love towards pictures of individuals with intellectual disabilities. Furthermore, the involvement of the striatum in feelings of love is substantiated by two studies: one in which participants saw a beloved person (Aron et al., 2005) and one in which subjects looked at smiling faces (Vrtička et al., 2008). Given that these regions are linked to reward processing and show a high density of oxytocin and vasopressin receptors – neuropeptides that play a crucial role in attachment and bonding (for review, see Depue & Morrone-Strupinsky, 2005; Zeki, 2007), the described results might be interpreted as reflecting the rewarding nature of experiencing love and warmth. In addition, recent findings suggest that middle insula activations may be linked to adopting a

compassionate attitude when witnessing the suffering of others. In a paradigm that involved listening to distressing sounds while being immersed in a compassionate state, expert meditators had more pronounced middle insula activations than novice meditators (Lutz et al., 2008).

Taken together, these results suggest that the previously introduced distinction between empathic distress and compassion as the two consequences of empathy is paralleled by different neural substrates. Whereas distressing empathic experiences have been shown to be associated with AI and ACC, compassionate or loving experiences seem to involve mid-insular and striatal regions. Due to the scarcity of research in this field, many more studies are needed to refine the delineation of the neural networks involved in the positive emotions of compassion and love and compare those to sharing the negative feelings such as pain or unpleasant tastes and odors.

Summary

The field of social neuroscience is evolving quickly and the above described studies have (among others) greatly advanced our understanding of the neural bases of empathy. Most importantly, the development and use of ecologically valid paradigms could show that self-experienced and vicariously experienced pain rely on shared neural substrates in the AI and ACC. Together with the finding of distinct activation patterns in self-experienced affective states, this suggests that we understand the feelings of others by simulating the affective component of the observed states. In this context, the insula plays a specific role, since it generally serves as an interoceptive cortex supporting representations and predictions of feeling states for oneself and for others. Despite the major advances in our understanding of the empathic brain, many questions remain to be answered while at the same time new questions arise with newly gained insights.

Outstanding Questions

- In light of the findings speaking to a shared network for self-experienced and vicariously experienced emotions, how far are the neural substrates shared on the level of single neurons? Can these phenomena be distinguished by functional gradients in structures like the insula?
- By which mechanisms can empathic experiences be influenced so that instead of resulting in empathic distress, empathy leads to compassion? Which neural changes accompany these interventions and how do neural pathways underlying compassion

differ from neural signatures of negative experiences associated with sharing someone else's pain?

- What role do neurotransmitters like oxytocin and vasopressin play in empathy and how do they interact in the circuitry of emotion processing?
- Given that the acquisition of empathic and compassionate skills probably depend on the maturation of certain cortical structures, which neural changes accompany the development of empathy and related concepts from early childhood to adolescence? How do these findings tie in with emotional plasticity over the lifespan?

Methods: Magnetic Resonance Imaging

The central aim of this thesis was to determine the neural plasticity underlying the training of compassion and empathy. To localize these short-term changes in the brain, we used blood oxygenation level dependent (BOLD) signals acquired with functional magnetic resonance imaging (fMRI). Another goal of this thesis was to identify structural connectivity patterns related to empathy. This was accomplished by using structural covariance analysis of *in vivo* cortical thickness measures.

Magnetic Resonance imaging (MRI)

In the following section, the basis of MRI will be briefly discussed. Detailed accounts of the underlying methods can be found elsewhere (Jäncke, 2005; Jezzard, Matthews, & Smith, 2001). MRI relies on the different magnetic properties of hydrogen (H) atoms in tissue water. Hydrogen nuclei consist of a single proton that randomly spins around its axis. When applying an external magnetic field, such as in the MR scanner, H nuclei can assume a high or low energy state. In high energy states, H nuclei behave as if they are oriented against the applied field and in low energy states they behave as if aligned with the magnetic field. Most H atoms assume a low energy state and align with the magnetic field. By applying electromagnetic pulses in the range of radiofrequency (RF), the nuclei can be excited into a high energy state. When subsequently returning into the low energy state, measurable RF energy is emitted. This relaxation signal decreases exponentially and is the basis of three different measures: T1, T2 and T2*:

T1 relaxation reflects the recovery of the longitudinal magnetization to equilibrium. In other words, the nuclei return from the high energy state in which the RF pulse excited them to the low energy state. This allows the assessment of anatomical structures, since T1 signals are different for cerebrospinal fluid (CSF), grey matter (GM) and white matter (WM). Key to this differentiation of neural structures are two aspects: firstly, the different distribution of protons in water and fat and secondly, characteristic relaxation times that arise due to the interaction of the nuclear spins with their environment (spin-lattice interactions). Contrasts can be obtained by applying the pulses faster than the different molecules would require to completely relax. For instance, due to the abundant amount of water, the signal in CSF decays very slowly, while signals in brain tissue decay much faster. Changes in the contrasts can be generated through varying the repetition time (TR).

T2 contrasts are based on transverse relaxation. Due to spin-spin interactions between a large number of nuclei, the induced longitudinal magnetization does not return to equilibrium homogeneously. In fact, local changes in magnetic fields at the molecular level allow the exchange of energy between nuclei. This leads to a loss in phase coherence and reduces the intensity of the emitted signal. In other words, T2 relaxation contrasts reflect the chemical properties of tissue types. Increased delays between the measurements of signals (time to echo, TE) augment the difference between signals from tissues with longer T2, like GM, compared to tissues with shorter T2, like WM.

Finally, T2* relaxation results from variations in the applied static magnetic field. These variations mainly arise from differences in the geometry and composition of the imaged sample. A prominent example is the boundary between tissue and air (as encountered at the sinuses), which can cause very steep differences in the magnetic field and lead to rapid T2* relaxation. Importantly, T2* signals are also susceptible to the different magnetic properties of oxygenated (diamagnetic) and deoxygenated (paramagnetic) blood. While oxygenated blood is diamagnetic (has reduced magnetic flux), deoxygenated blood is paramagnetic (has increased magnetic flux). Since the ratio of oxy- to deoxyheamoglobin is related to neural activity, T2* weighted contrasts allow non-invasive *in vivo* measures of brain function to be obtained.

Functional Magnetic Resonance Imaging (fMRI)

BOLD responses that rely on T2* contrasts can serve as an indirect measure of changes in the spatio-temporal distribution of neural function during specific tasks or states. Experiments comparing fMRI signals to electrical neural activity determined that BOLD signals are roughly equivalent to local field potentials (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001). The biological assumption of the BOLD signal is that increases in neural activity rely on augmented metabolic processes, such as synaptic activity. Through the consumption of oxygen, metabolic processes cause an initial dip in signal intensity (see Figure 1.3). This deoxygenation leads to higher demands of oxygenated blood in the activated neural region. As the requirement of oxyhemoglobin is lower than the amount of oxygenated blood provided, there is a surplus of oxygenated blood in the activated brain region. Owing to its diamagnetic properties, oxygenated blood decreases field distortions around the vessels and thereby increases the local signal. This typically occurs over 2–4 s after stimulus presentation. The signal then remains stable over some time and typically decays over 12–18 s, followed by a post-stimulus undershoot.

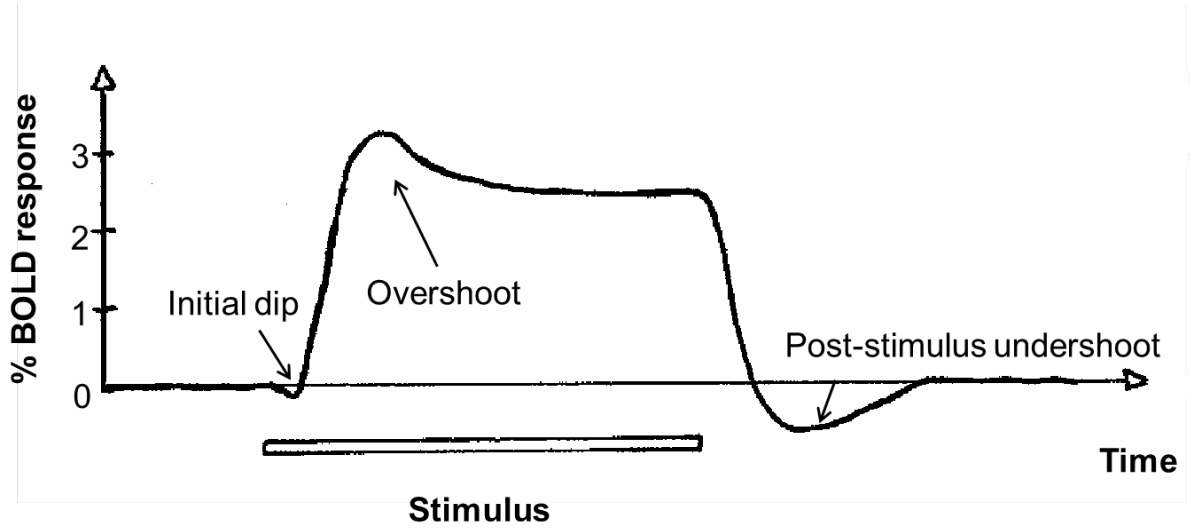


Figure 1.3. Time course of a typical hemodynamic response.

Structural Covariance Analysis and Cortical Thickness Measurements

The following section will briefly introduce the basis of MRI-based cortical thickness measurements and the framework of covariance analysis for mapping structural networks *in vivo*. Previous methodological accounts have provided more detailed descriptions of cortical thickness measurements (Fischl & Dale, 2000; J. S. Kim et al., 2005; Lerch & Evans, 2005; MacDonald, Kabani, Avis, & Evans, 2000) and described the framework of structural covariance network mapping in more detail (for review e.g., Bullmore & Sporns, 2009).

The measurement of cortical thickness is based on structural brain images that are generally acquired using high-resolution T1-weighted imaging sequences. Although they differ with respect to several implementation details, the two most widely used software packages, FreeSurfer and CIVET, follow a similar order of steps to generate these measures. In brief, images undergo an initial preprocessing that involves intensity normalization and registration to a standard stereotaxic space. Subsequently, images are automatically classified into the tissue types GM, WM, and CSF. As depicted in Figure 1.4, anatomical surface-models are then fitted to follow the inner cortical boundary (i.e., between WM and GM) and the outer cortical boundary (i.e., between GM and CSF) using deformation algorithms. Cortical thickness is measured as the distance between corresponding points of the inner and outer cortical surface. To improve the correspondence of measurement points across individuals, surfaces are aligned with

respect to cortical folding patterns. In a separate step, surface-based diffusion smoothing may be used to reduce measurement noise and to increase statistical sensitivity.

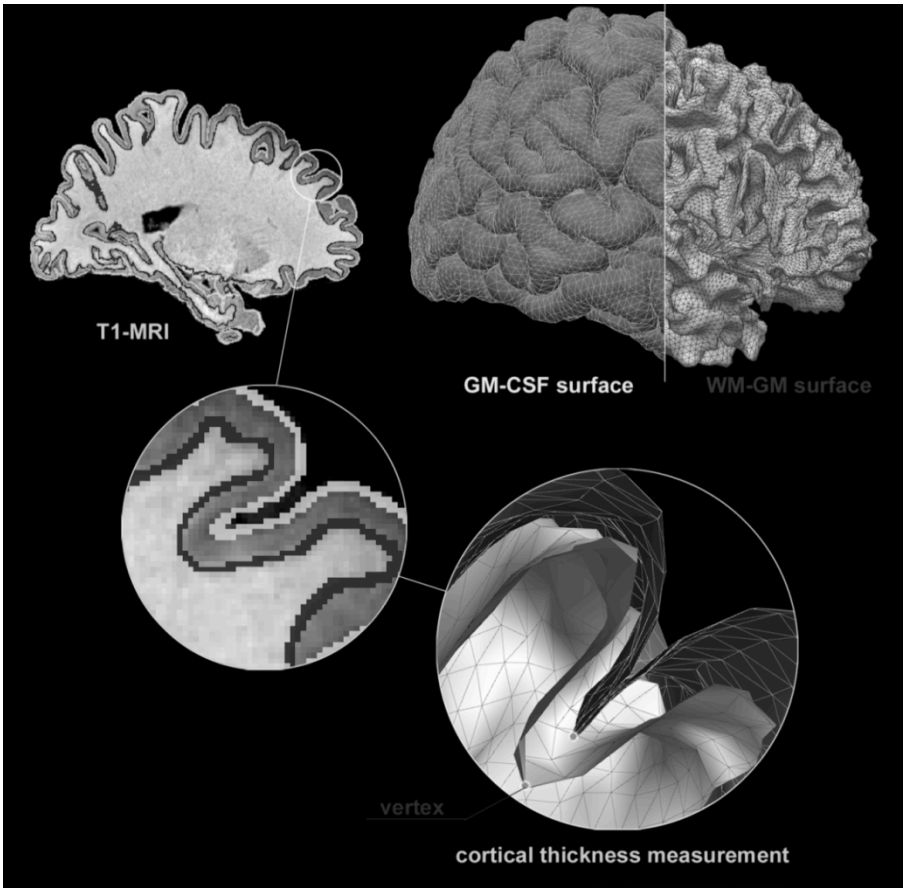


Figure 1.4. Cortical thickness is measured as the straight-line distance between the inner cortical boundary (WM-GM surface) and the outer cortical boundary (GM-CSF surface). With kind permission from Boris Bernhardt.

By respecting the topology of the cortical mantle, such cortical thickness measures offer a crucial advantage over automated structural MRI analysis using voxel-based morphometry (VBM), a technique that statistically compares smoothed classification maps in standard space (for discussions, see Ashburner & Friston, 2001; Bookstein, 2001; Winkler et al., 2012). The neglect of anatomically important relationships in VBM is a potential source of erroneous inference. For example, volume-based registration, as performed during the preprocessing pipeline of VBM, may not account for the anatomical variability manifested in complex gyrification and sulcation patterns across individuals. This may reduce the specificity and

sensitivity with which significant effects can be detected. Moreover, voxel-wise isotropic smoothing averages regions that are close in voxel-space, but far apart on the convoluted cortical surface, such as adjacent gyral crowns. Voxel-based smoothing, furthermore, increases the amount of unwanted partial volume effects (the blurring between two tissue types), especially at tissue boundaries. Lastly, the interpretation of VBM-based statistical analyses can be difficult, as this measure likely reflects a complex combination of thickness alterations, cortical folding differences, and registration error.

We, thus, decided to perform structural covariance analyses based on cortical thickness measures. Under the assumption that the correlations of morphological measures between different brain regions reflect a form of inter-regional coupling, this method allows the non-invasive *in vivo* mapping of structural networks (Mechelli, Friston, Frackowiak, & Price, 2005). Correlational measures do not necessarily indicate the existence of a physical connection, such as a fiber path, between two regions (Gong, He, Chen, & Evans, 2012). However, inter-regional coupling in structural characteristics across subjects most likely reflects persistent functional-trophic interactions, and common genetic and developmental influences (Raznahan et al., 2011; Zielinski, Gennatas, Zhou, & Seeley, 2010). Furthermore, recent reports have shown a relatively strong convergence between structural covariance patterns and networks obtained from resting-state functional MRI data (e.g., Kelly et al., 2012; Seeley, Crawford, Zhou, Miller, & Greicius, 2009), speaking to this method's external validity.

2. Short-Term Compassion Training Increases Prosocial Behavior in a Newly Developed Prosocial Game

Corresponding publication:

Leiberg, S., Klimecki, O., & Singer, T. (2011). Short-term compassion training increases prosocial behavior in a newly developed prosocial game. *PLoS One*, 6(3).

Abstract

Compassion has been suggested to be a strong motivator for prosocial behavior. While research has demonstrated that compassion training has positive effects on mood and health, we do not know whether it also leads to increases in prosocial behavior. We addressed this question in two experiments. In Experiment 1, we introduce a new prosocial game, the Zurich Prosocial Game (ZPG), which allows for repeated, ecologically valid assessment of prosocial behavior and is sensitive to the influence of reciprocity, helping cost, and distress cues on helping behavior. Experiment 2 shows that helping behavior in the ZPG increased in participants who had received short-term compassion training, but not in participants who had received short-term memory training. Interindividual differences in practice duration were specifically related to changes in the amount of helping under no-reciprocity conditions. Our results provide first evidence for the positive impact of short-term compassion training on prosocial behavior towards strangers in a training-unrelated task.

Prosocial behavior is a pervasive aspect of human life: We cooperate with others and help them when they are in need. However, diametrically opposed to these behaviors are everyday experiences of people taking advantage of others. The present study is concerned with the question whether compassion training can increase prosocial behavior. Compassion has been defined as the emotion one experiences when one feels concern for another's suffering and desires to enhance that person's welfare (Baumeister & Vohs, 2007; see Singer & Steinbeis, 2009; Fehr, Sprecher, & Underwood, 2009 for more detailed definitions). In the present paper, we use the term ''compassion'' to describe an emotional as well as a motivational state, characterized by feelings of warmth, love, and concern for the other as well as the desire to help and promote the other's welfare. The term ''empathic concern'' has been used in a very similar way in developmental and social psychology (Davis, 1983; Batson, 2009). For example, Batson (Batson, Duncan, Ackerman, Buckley, & Birch, 1981) maintains that empathic concern ''is an other- oriented emotional response elicited by and congruent with the perceived welfare of someone in need involving feelings for the other such as sympathy, compassion, tenderness, and the like.'' However, while empathic concern mainly denotes a situation- specific, rather short-living emotion, compassion can also be thought of as an attitude (Gilbert, 2010). Empirical evidence suggests that empathic concern is a perpetuator of prosocial behavior (Batson, 1991; Eisenberg & Fabes, 1998). For example, it has been demonstrated (Batson, Eklund, Chermok, Hoyt, & Ortiz, 2007) that momentarily inducing feelings of empathic concern for a person in need by having participants focus on the person's feelings increases their prosocial behavior towards that person. More specifically, participants who were instructed to feel empathic concern for a person receiving painful electric shocks were willing to receive more shocks themselves to alleviate the other person's suffering than participants who had been encouraged to remain detached. The effects of this situation-specific induction of empathic concern, however, are probably rather short-lived and might not extend over the particular experimental session. Furthermore, it is not clear whether the induction of empathic concern for a specific person leads to increases of prosocial behavior only for that specific person (e.g. Batson et al., 1981) or whether it generalizes to different persons as well (Greitemeyer, 2009; Twenge, Baumeister, DeWall, Ciarocco, & Bartels, 2007). While the experimental induction of empathic concern through explicit perspective-taking instructions or listening to songs with prosocial lyrics (Greitemeyer, 2009) might temporarily prime people to experience empathy when seeing the distress of others, training of compassion aims at permanently changing people's motivation and their feelings towards other people. It strives to develop a more friendly, benevolent, connected and positive attitude towards others. In the long run, compassion training-induced

changes at the trait level – but not at the state level - might even take effect on the opiate- and oxytocin-based affiliative system (Gilbert, 2010; Depue & Morrone-Strupinsky, 2005).

We hypothesize that, contrary to a short-term instruction-based induction of empathic concern towards a specific person, compassion training will elicit a longer-lasting enhancement of general compassionate motivation, which in turn may lead to an increase in the general tendency to act prosocially, independent of person and situation.

Even though there is a long history of behavioral plasticity research pertaining to the training of cognitive (Lustig, Shah, Seidler, & Reuter-Lorenz, 2009), perceptual (Fahle, 2005), motor (Gentili, Han, Schweighofer, & Papaxanthis, 2010) as well as affective skills (Golan et al., 2010; Slaski & Cartwright, 2003), no study to our knowledge has investigated behavioral changes resulting from compassion training. Thus, in one study, for example, empathy for a personal offender was trained over eight 1-hour sessions and an increase in reported empathy and forgiveness but not prosocial behavior was measured (McCullough, Worthington, & Rachal, 1997). Similarly, the few existing studies on compassion training have examined the effects of compassion training on mood and health but not prosocial behavior (Carson et al., 2005; Fredrickson, Cohn, Coffey, Pek, & Finkel, 2008; Pace et al., 2009). In a pilot study, Gilbert and Procter (Gilbert & Procter, 2006) administered compassionate mind training (CMT; Gilbert & Irons, 2005), which aims at reducing self-criticism by focusing on compassionate images and emotions,

to a sample of psychiatric patients with severe long-term difficulties. They reported reductions in depression and anxiety as well as increases in self-soothing abilities and feelings of warmth for oneself. Other studies investigating the effects of compassion training have used meditation-based techniques that involve the development of warm, positive feelings towards a variety of people and ultimately towards all human beings: Six to seven weeks of meditation-based compassion training result in increases in positive mood and life satisfaction (Fredrickson et al., 2008) as well as a reduction of interleukin-6 release in response to a psychosocial stressor (Pace et al., 2009). The more time participants had actually spent training, the stronger the reduction in interleukin-6 release, suggesting a dose- dependent effect of compassion training. Hutcherson et al. (Hutcherson, Seppala, & Gross, 2008) report that a very brief (7-min) compassion meditation exercise results in a more positive attitude towards the target of the exercise. Taken together, these studies provide promising support for the health- and positivity-promoting effects of compassion training. However, so far, no study has investigated whether

prosocial behavior can actually be increased through compassion training and whether the practice of compassion promotes a generalized tendency for prosocial behavior. Thus, the aim of the present study was to investigate the effect of short-term compassion training on prosocial behavior.

In behavioral economics, prosocial behavior is usually studied in the context of well-controlled monetary exchange games (Camerer, 2003) and mostly explained in terms of social preferences or norms, such as fairness and reciprocity (Fehr & Schmidt, 2006), whose evolution has also been linked to reputation concerns (Nowak, Page, & Sigmund, 2000; Nowak & Sigmund, 2005). It could also be shown that observation of prosocial behavior in a public goods game with multiple rounds increases the likelihood of later prosocial behavior of the observer towards another person in the following rounds (Fowler & Christakis, 2010). However, the influence of compassion or empathy and their training on prosocial behavior has so far never been discussed or studied in the field of economics. In the context of game theoretical paradigms, the dictator game is most commonly used for assessing altruistic acts towards others (Hoffman, McCabe, Shachat, & Smith, 1994; Bohnet & Frey, 1999; Eckel & Grossman, 1996; Forsythe, Horowitz, Savin, & Sefton, 1994). In this game, participants are endowed with a sum of money that they can split between themselves and another participant who has no money. Giving in the dictator game is likely driven by fairness norms and not by kindness (Bolton, Katok, & Zwick, 1998). While several motives have been discussed as underlying prosocial behavior, only recently a differentiation between norm-based and compassion-based prosocial behavior has been suggested (Singer & Steinbeis, 2009). While the former is particularly encountered in ‘‘cold’’, reasoning-driven exchange situation, the latter is often present in ‘‘hot’’, emotion-provoking situations. Compassion training might take its effects on the latter but not the former. Since many of our everyday interactions are not purely rational, but involve emotions, an adequate paradigm that assesses prosocial behavior in an engaging, ecological setting and that is sensitive to affective interventions needs to be developed. This paradigm would moreover allow for future investigation of the proposed differentiation between norm-based and compassion based prosocial behavior.

In social psychology, prosocial behavior is mostly assessed in emotion-provoking one-shot helping situations of high ecological validity, such as dropping pens, soliciting donations for charities, or soliciting help with filling out or scoring questionnaires (Batson et al., 1981; Van Lange, Bekkers, Schuyt, & Vugt, 2007; Bartlett & DeSteno, 2006; Levine, Martinez, Brase, & Sorenson, 1994). However, these paradigms as well as the above-mentioned economic

paradigms do not allow for the repeated assessment of prosocial behavior within the same person, which is required in intervention studies with multiple measurement time points such as the present study. We therefore developed a new prosocial task – the Zurich Prosocial Game (ZPG) – that allows for the repeated assessment of prosocial behavior within the same person while still being ecologically valid, and thus being suitable to investigate changes in prosocial behavior due to compassion training.

In addition the new game was developed to simultaneously assess the influence of reciprocity, the cost associated with helping, and distress cues on prosocial behavior. It has been shown that people help more often if they have been helped before (Falk & Fischbacher, 2006; Wilke & Lanzetta, 1970), if the costs of helping are low (Dovidio, Piliavin, Gaertner, Schroeder, & Clark, 1991) and if they are confronted with signs of distress (Batson & Coke, 1981; Carlo, Eisenberg, Troyer, Switzer, & Speer, 1991). These factors are of interest as evolutionary biologists and anthropologists demonstrated that they are selected for in evolution and provide a biological basis for altruism. Reciprocal altruism evolved as a costly altruistic act which might be repaid at a later time (Trivers, 1971), costly helping is mostly directed towards kin as suggested by the model of inclusive fitness (Burnstein, Crandall, & Kitayama, 1994) and distress cues, such as crying, evolved to signal the need for help and to sustain close personal bonds (Nelson, 2005). The possibility to distinguish between these helping-related factors within one task allows the investigation of differential effects of context, intervention or personality on different helping conditions in future research. Here, the aim was first to test the effect of a short-term compassion training on prosocial behavior in the ZPG.

To validate the newly developed task and to test the effects of compassion training on prosocial behavior, we performed two independent experiments. The first experiment was conducted to validate the newly developed prosocial task, the so-called Zurich Prosocial Game (ZPG) and to test its sensitivity to the influence of reciprocity norms, helping costs and distress cues on helping. We hypothesized that people would help more a) if they had been helped before, b) if the cost of helping was low, and c) if they were confronted with distress cues. The second experiment was conducted to investigate the influences of short-term compassion training on prosocial behavior towards strangers as measured by the ZPG – a game that is completely unrelated to the training context. We hypothesized that short-term compassion training leads to stronger increases in helping than a short-term memory training, the latter received by a control group. Furthermore, time spent practicing the compassion-enhancing technique should be positively correlated with this increase in helping. Based on the suggested distinction between

compassion-based and norm-based prosocial behavior, and on the assumption that the compassion training has effects on the former this correlation could possibly only arise for non-reciprocity trials.

Results

Experiment 1

To investigate the effects of reciprocity, cost, and distress on the occurrence of prosocial behavior, we computed a 2 (reciprocity: no reciprocity, reciprocity) x 2 (cost: low, high) x 2 (distress: no distress cues, distress cues) within-subjects repeated-measures analysis of variance (ANOVA). This analysis revealed main effects of reciprocity, cost, and distress (see Table 2.1). As hypothesized, participants helped significantly more in reciprocity trials, in trials with a low cost of helping, and in trials in which the co-player's virtual character expressed distress (see Materials and Methods for a detailed description of the ZPG). Thus, the ZPG indeed seems to be sensitive to the three operationalizations of the influencing factors, which suggests that reciprocity, distress cues, and low cost are associated with increased helping behavior (see Figure 2.1).

The main effects of reciprocity and cost were qualified by a significant interaction between the two factors (see Table 2.1). Increasing the cost of helping resulted in a larger decrease in prosocial behavior in the no-reciprocity compared to the reciprocity trials, suggesting that norms such as reciprocity can absorb the decline in prosocial behavior when helping is costly. Furthermore, there was a three-way interaction between the three factors (see Table 2.1). Separate follow-up ANOVAs for distress and no-distress trials revealed that the reciprocity x cost interaction was only significant in the distress trials, $F(1, 67) = 17.73$, $p < .001$, partial $\eta^2 = .21$.

To confirm that inter-individual differences in helping behavior as measured using the ZPG are not brought about by differences in allocation of attention, we calculated the correlation between the total amount of helping and the percentage of stars picked up, that randomly appeared during the game. On average, participants picked up the star in 42.6% of the trials in which a star appeared. A star appeared in two to six (of nine) trials. The non- significant correlation, $\rho(66) = -.07$, $p > .05$, between total amount of helping and percentage of stars picked up indicates that

differences in attention allocation most likely do not account for inter-individual differences in prosocial behavior.

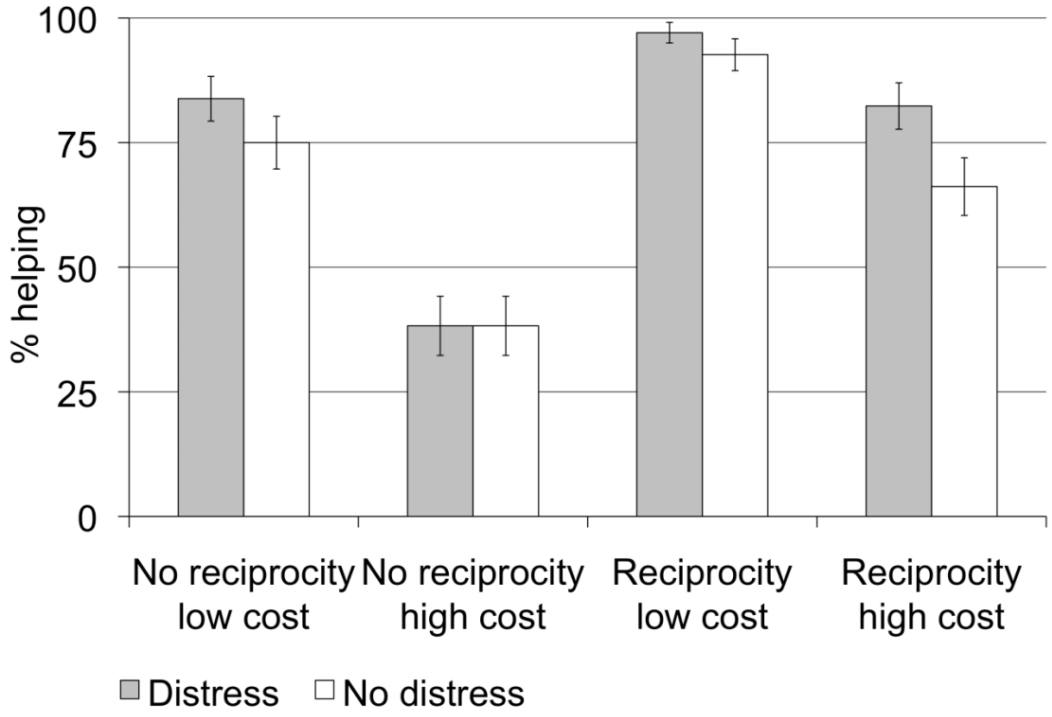


Figure 2.1. Percent helping in the different conditions of the ZPG. Error bars denote standard errors of mean.

Furthermore, to control for potential effects of individual differences in risk preferences on helping in the high cost trials, the risk questionnaire and the lottery index were correlated with helping in high cost trials (see Materials and Methods for a description of the risk perception control measures). As both correlations were non-significant, we can preclude that inter-individual differences in risk preferences, $\rho(66) = .10$, $p > .05$ (risk questionnaire) and $\rho(66) = .02$, $p > .05$ (lottery index), accounted for the difference in prosocial behavior between low and high cost trials.

Table 2.1

Source	df	*F*	partial η^2	*p*
Experiment 1				
Reciprocity	67	73.22	.52	< .001
Cost	67	73.78	.52	< .001
Distress	67	7.02	.10	= .01
Reciprocity x Cost	67	13.21	.17	= .001
Reciprocity x Cost x Distress	67	4.10	.06	< .05
Experiment 2				
Reciprocity	68	51.55	.43	< .001
Cost	68	66.04	.49	< .001
Reciprocity x Cost	68	7.96	.11	< .001
Reciprocity x Distress	68	6.11	.08	< .05

Note. ANOVA for the effects of reciprocity, cost and distress cues on prosocial behavior in experiment 1 and 2. All main effects and interactions significant on a $p < .05$ level are reported.

Furthermore, participants had to judge their engagement in the game (see Materials and Methods). The analyses of these subjective engagement scores revealed that on average participants indicated that they were very engaged in the game (range: 2–5; mean = 4.15, SD = 0.74); a result which matches the observation of the experimenter who reported that the subjects were all very immersed in the ZPG.

To assess the convergent validity of the ZPG, participants played the dictator game (Hoffman et al., 1994; see Materials and Methods). Based on our reasoning about norm-based and compassion-based prosocial behavior in the introduction, we did not expect an exceedingly high

correlation between the ZPG and the dictator game, but still, as both tasks assess variants of prosocial behavior, a sufficiently high correlation to maintain that the ZPG indeed measures prosocial behavior. In the dictator game, participants gave 36.29% of their endowment on average. Most of the participants (40.3%) gave half of their endowment and 10.4% gave nothing. As expected, giving behavior in the dictator game correlated with helping behavior in the ZPG, $\rho(65) = .35$, $p = .004$, substantiating the validity of our game as a measure of prosocial behavior.

To assess the divergent validity of the ZPG, we used a memory task (see Materials and Methods). Participants remembered 18.24 words (SD = 5.86) on average in the memory task. The number of remembered words did not correlate significantly with helping in the ZPG, $\rho(64) = .06$, $p > .05$, demonstrating divergent validity of the ZPG.

Experiment 2

In the following, we will first present the pre-training data from the newly developed ZPG to ascertain whether the results found in Experiment 1 are robust. We will then report data on the effectiveness of the compassion training workshop and on the effects of compassion training on prosocial behavior in the ZPG. The effects of compassion training were tested one-sided as we had clear hypotheses about the direction of effects (see Introduction).

Robustness of ZPG. To investigate the robustness of the result pattern in the ZPG, we computed a 2 (reciprocity: no reciprocity, reciprocity) x 2 (cost: low, high) x 2 (distress: no distress cues, distress cues) within-subjects repeated-measures ANOVA for the total sample of Experiment 2 (compassion group and memory group). The analysis again revealed main effects of reciprocity, and cost (see Table 2.1). Participants helped significantly more in reciprocity trials and in trials with a low cost of helping. As in Experiment 1, these main effects were qualified by a significant interaction between the two factors, again suggesting that norms such as reciprocity can absorb the decline in prosocial behavior when helping is costly. In contrast to Experiment 1, however, no main effect of distress was observed, $F(1,68) = 1.89$, $p < .05$, partial $\eta^2 = .03$. The analysis did however yield a reciprocity x distress interaction: Distress cues increased helping in the no-reciprocity but not in the reciprocity trials, which might indicate that no- reciprocity trials are more sensitive to other influencing factors (see Table 2.1).

We again did not observe a significant correlation between total amount of helping and percent of stars picked up, $\rho(67) = -.12$, $p > .05$, indicating that differences in attention allocation most likely do not account for inter-individual differences in helping behavior.

As in Experiment 1 the correlations between helping in the high cost trials with both the risk questionnaire, $\rho(57) = .02$, $p > .05$, and the lottery index, $\rho(57) = .05$, $p > .05$, were non-significant, precluding that inter-individual differences in risk preferences accounted for the difference in prosocial behavior between low and high cost trials.

Participants in Experiment 2 also reported to be very engaged in the game (range: 1–5; Mean = 4.04; SD = 1.07). There was no difference between participants in the compassion and memory training group in the engagement with the game, $t(52) = 1.78$, $p > .05$, suggesting that potential differences between the groups cannot be accounted for by differences in motivation and degree of being emerged into the game.

In the dictator game, on average, participants gave 33.8% of their endowment. Most of the participants (39.4%) gave half of their endowment and 13.6% gave nothing. More importantly, giving behavior as measured with the dictator game again correlated with helping behavior as measured with the ZPG, $\rho(67) = .45$, $p < .001$.

Participants remembered on average 20.53 words (SD = 7.04) in the memory task. And as in Experiment 1, the number of remembered words did not correlate with helping in the ZPG, $\rho(66) = -.04$, $p > .05$, giving repeated evidence for divergent validity of the ZPG.

Effectiveness of compassion training. Repeated-measures ANOVAs with time (pre-training, post-training) as a within- subjects factor and training (compassion, memory) as a between-subjects factor were calculated to determine the effectiveness of the compassion training in enhancing self-reported positive (assessed with the Positive and Negative Affect Scale; PANAS; Watson, Clark, & Tellegen, 1988) and compassionate (assessed with the Compassionate Love Scale; CLS; Sprecher & Fehr, 2005) feelings and reducing negative feelings (also assessed with the PANAS; Watson et al., 1988). A significant main effect of time on positive mood, $F(1,54) = 23.47$, $p < .001$, partial $\eta^2 = .30$, was revealed, indicating that compassion training as well as memory training increased positive mood. A significant main effect of time, $F(1,54) = 5.84$, $p = .02$, partial $\eta^2 = .10$, was revealed for compassionate feelings that was qualified by a marginally significant interaction between time and training, $F(1,54) = 3.61$, $p = .06$, partial $\eta^2 = .06$. Post-hoc t-tests indicated that only the compassion-training group experienced a

significant increase in compassionate feelings, $t(23) = 2.66$, $p = .01$. For negative mood, a significant time x training interaction was revealed, $F(1,68) = 6.11$, $p = .016$, partial $\eta^2 = .08$. While negative mood decreased in the compassion-training group, $t(23) = -1.94$, $p = .03$, one-sided, it marginally significantly increased in the memory-training group, $t(23) = 2.02$, $p = .05$.

Effect of compassion training on prosocial behavior. To test whether a brief compassion training had an effect on prosocial behavior in the ZPG, we conducted two analyses: First, we performed a 2 (time: pre-training, post-training) x 2 (reciprocity: no reciprocity, reciprocity) x 2 (cost: low, high) x 2 (distress: no distress cues, distress) repeated-measures ANOVA with training (compassion, memory) as a between-subjects factor. Second, we tested for increases in helping as a function of inter-individual differences in hours of reported training. To this end, we calculated the correlation between participants' self-reported time spent practicing outside of the training and the change in helping from pre- to post-training (self-report data could only be obtained from a subset of the samples: n(compassion) = 19, n(memory) = 22).

In the first analysis, we observed a significant time x training interaction, $F(1,57) = 4.09$, $p = .05$, partial $\eta^2 = .07$. While there was no reliable change in helping from pre- to post-training for the memory training group, $t(31) = -1.20$, $p = .24$, compassion training significantly increased helping, $t(26) = 1.85$, $p = .04$, one- sided (see Figure 2.2). Additionally, a time x cost interaction was observed, $F(1,57) = 6.76$, $p = .01$, partial $\eta^2 = .11$. These interactions were qualified by a significant three-way interaction between time x cost x training, $F(1,57) = 4.55$, $p = .04$, partial $\eta^2 = .07$. Follow-up independent t-tests indicated that, at pre-training, helping in the low-cost, $t(57) = 1.21$, $p = .23$, and high-cost trials, $t(57) = 0.55$, $p = .58$, did not differ between the compassion and the memory group whereas, at post-training, the groups differed significantly in helping in both the low-, $t(57) = 3.07$, $p = .003$, and the high-cost trials, $t(57) = 2.27$, $p = .03$.

The second analysis did not reveal the hypothesized correlation between inter-individual differences in reported hours of compassion training and total helping, $\rho(17) = .27$, $p = .13$, one-sided. To test our expectation that inter-individual differences in reported training hours are differentially related to the different trial types, affecting more non-reciprocity than reciprocity trials, we calculated the correlation between training hours and helping in the different trial types and found a significant correlation between inter-individual differences in reported hours of compassion training and helping in no-reciprocity trials, $\rho(17) = .39$, $p = .05$, one-sided. Correlations with the other trial types as well as all correlations in the memory group were non-significant at $p < .05$, one-sided.

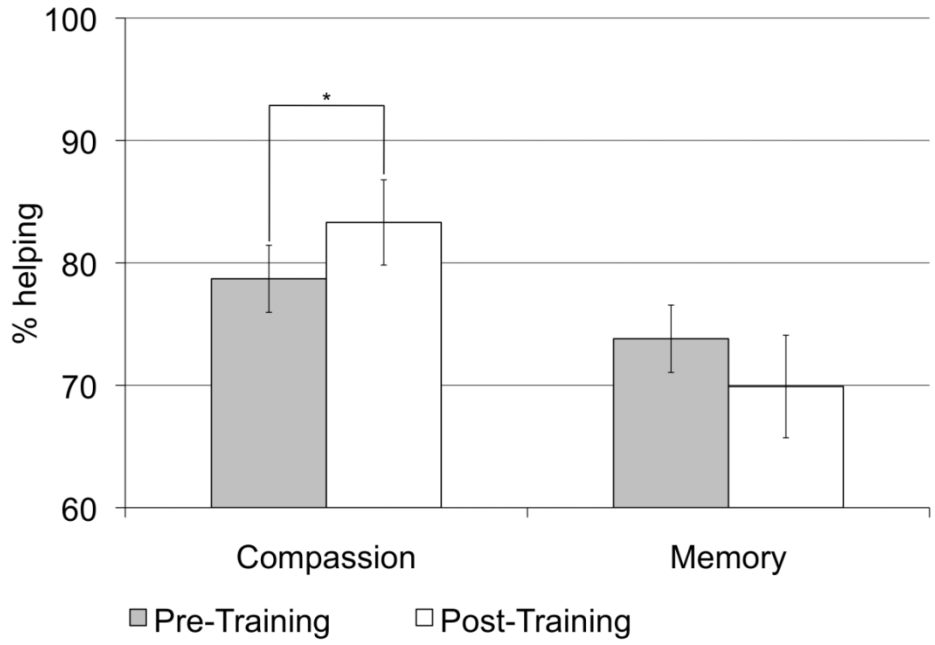


Figure 2.2. Effects of training on overall helping in the ZPG for the compassion-training and memory-training group. Error bars denote standard errors of mean. * $p < .05$, one-sided.

To investigate whether compassion training could also increase giving in the dictator game, we computed a repeated-measures ANOVA with time (pre-training, post-training) as a within-subjects factor and training (compassion, memory) as a between- subjects factor. Interestingly, neither a significant main effect of time, $F(1,56) = 0.02$, $p = .89$, partial $\eta^2 < .001$, nor an interaction between time and training $F(1,56) = 1.63$, $p = .21$, partial $\eta^2 < .001$, was observed. Participants in neither group gave more money to the other person after training compared to before training.

Discussion

The present study introduces a newly developed prosocial game – the Zurich Prosocial Game (ZPG) – and provides first evidence for the effectiveness of short-term compassion training in enhancing prosocial behavior in this new training-unrelated game towards strangers.

The ZPG was developed as previous prosocial tasks from behavioral economics or social psychology are either not particularly ecologically valid or do not allow for the repeated assessment of prosocial behavior which is required in intervention studies with multiple measurement time points. The ZPG extends the prosocial tasks from behavioral economics and social psychology in several aspects. First, the influence of reciprocity, cost, and distress on prosocial behavior has been studied separately before, but the ZPG now has the advantage to allow their simultaneous assessment in the same setting. This is particularly useful when studying the differential influence of experimental manipulations on these factors. Second, while many prosocial tasks are only applicable one time, the ZPG can be played multiple rounds and on different time points with the same subjects thus allowing for more stable estimates of prosocial behavior, the usage in neuroscientific settings where many trial repetitions are needed, and, the assessment of changes in prosocial behavior over time through interventions. Third, the ZPG is more ecologically valid than, for example, monetary exchange games, as it minimizes the influence of strategic considerations, minimizes effects of task- affordances due to explicit instructions and maximizes the influence of emotion-driven, fast decisions, since participants are immersed in the game itself whose explicit goal is to achieve a treasure in short time rather than act prosocially. In the ZPG participants help others by spending resources (key, time) they might need later on. This type of prosocial behavior that involves uncertainty for oneself can be encountered often in daily life, for example when we run for an important appointment and see someone fall from his bike. Do we stop to help this person without knowing the outcome and how much time it will take or do we refrain from helping and make sure that we reach our appointment on time? And fourth, as the ZPG is very engaging and easy to use, it is also very well suited to study prosocial behavior in children.

The present results confirm that the ZPG is sensitive to influences of reciprocity, cost, and distress on prosocial behavior: As predicted, participants of two independent experiments helped more when having been helped before and when costs are low. Interestingly, the drop in prosocial behavior with increasing costs was less pronounced when participants had been helped before suggesting that norms of reciprocity override cost considerations. While in Experiment 1 participants helped more when confronted with distress cues, in Experiment 2, distress cues increased helping only in the no-reciprocity trials but not in the reciprocity trials.

This may again suggest that reciprocity norms are so pervasive that they overrule the effect of any other influencing factor, whereas prosocial behavior without reciprocation is more affected by other factors. In both samples, the convergent validity with a well-established economic

prosocial task, the dictator game, was confirmed. This supports our claim that the new game does indeed assess prosocial behavior. The correlation between the two tasks, however, is modest, suggesting that the two measures tap into different aspects of prosocial behavior. Furthermore, divergent validity was established with a memory task.

In Experiment 2, we were able to demonstrate that compassion training but not memory training significantly increased helping in the ZPG. Previous studies have demonstrated that a momentary instruction-based induction of empathic concern for a specific person increases prosocial behavior towards that person immediately after induction (Batson, 1991). Here we show for the first time that compassion training had longer-lasting effects on prosocial behavior as the post-test was completed two to five days after training. Furthermore, short-term training resulted in transfer to behavior in a novel task that was completely unrelated to the previous affective training. Finally, compassion training increased prosocial behavior towards people who were not specifically targeted during training but complete strangers to the participants.

The present results support the notion that similar to situation-specific induction of empathic concern for a specific person in need (Batson, 1991; Eisenberg & Miller, 1987), the training of compassionate motivation leads to increases in prosocial behavior. In comparison to experimental inductions of empathic concern, however, compassion training has the potential to lead to longer-lasting changes in people's attitude and behavior towards other people that transcend the specific situation in which compassionate feelings were evoked and transfer to a much broader range of people and situations.

Self-benefiting effects of compassion training such as increases in positive mood, life satisfaction, decreased depressive symptoms (Fredrickson et al., 2008), and less reactivity to psychosocial stress (Pace et al., 2009) have been reported before. The present study adds to these findings by showing that even a short-term compassion training may not only have benefits for the practitioner's health and subjective well-being but also for other people and society in general as it increases the propensity to act prosocially even towards people one has never met. Notably, the prosocial behavior observed here was not directed towards a target of the compassion training but to random strangers and was assessed at least two days after the training, which lends further credibility to the societal impact that the implementation of compassion training in schools, organizations, and clinical settings might have (for compassion training in psychotherapy, see Gilbert, 2010).

Another interesting finding of the present study was that helping in no-reciprocity trials, but not helping in reciprocity trials, was related to inter-individual differences in reported training hours in the compassion group. This might provide tentative evidence for a differentiation between compassion-based and norm-based prosocial behavior as has been suggested before (Singer & Steinbeis, 2009). Accordingly, helping after having been helped may rely on a felt obligation to reciprocate cooperation. In contrast, helping without the possibility for reciprocity may be motivated more by feelings of compassion than by ‘‘cold’’ norms. The pattern of correlations found here suggests that compassion training might have differential effects on both types of underlying motivation. This is further supported by our finding that giving in the dictator game did not change from pre- to post-training in either group and that the modal giving at pre-training was 50%. Moreover, the correlation between helping in the ZPG and giving in the dictator was modest suggesting that the two measures tap into different aspects of prosocial behavior. Giving in the dictator game has previously been shown to depend more on fairness norms than on kindness (Bolton, Katok, & Zwick, 1998). These findings suggest a distinction between compassion-based and norm-based prosocial behavior with compassion training possibly exerting a stronger effect on the former than on the latter. As the current study was not designed to test the hypothesis of a distinction between compassion-based and norm-based prosocial behavior, future investigations are needed. For example, using priming of a reciprocity-norm or compassion could be used to show a differential effect of these concepts on different helping settings. The simultaneous assessment of reciprocity and non-reciprocity driven prosocial behavior in the ZPG makes this game ideal for this aim. Similarly, investigations with longer training will be of great interest. As the results suggest, the novel ZPG might be a better measure for assessing training- induced changes in prosocial behavior, specifically compassion- based prosocial behavior, than standard economic games or psychological measures as we were able to show that it is more sensitive to change than, for example, the dictator game. The higher sensitivity to changes in compassion-based prosocial behavior might result from the high emotional engagement participants experience when playing the ZPG.

Importantly, as the game is framed as a treasure hunt with monetary gains, demand effects induced by the content of the training should be less strong than in other prosocial tasks. While in economic games the sharing purpose is made explicit, here the instruction focuses on the rules of the game and emphasize that the goal is to reach the treasure in a limited time while having to overcome certain obstacles. Furthermore, the game is very engaging (on average participants rate their involvement in the game with 4 on a 1-to-5-scale and report later that they

find the game very enjoyable) and fast, thus making strategic considerations difficult. Compassion training not only increased prosocial behavior but also led to increases in reported compassionate feelings and positive affect and a decrease in negative affect. Interestingly, the memory-training group also evinced an increase in positive mood, suggesting that increases in positive mood are not sufficient for explaining enhanced prosocial behavior. We maintain that compassion training enhanced prosocial behavior through initial changes to participants' way of feeling and thinking about other people to a more positive, benevolent and friendly attitude. This is in line with participants' qualitative post-study reports of being more sensitive to others, feeling more connected, secure and open and having ''a bigger and more open heart.'' The present study provides first evidence for compassion training but not memory training causing increases in prosocial behavior. Future studies should elucidate, which aspects of the training led to the observed effect. Apart from the suggested change in other- related attitudes, increased relaxation or feeling of oneness (perceived self-other overlap; Cialdini, Brown, Lewis, Luce, & Neuberg, 1997) could be additional mechanisms through which compassion training increases prosocial behavior.

In sum, the present study provides first evidence for the effectiveness of a short compassion training in increasing prosocial behavior in a newly developed computer task, the Zurich Prosocial Game. Using this novel training-unrelated computer task, we found that compassion training that aimed at fostering a friendly, benevolent attitude towards others produced a significant increase in prosocial behavior two to five days after training towards strangers. Interestingly, practicing compassion strategies seems to influence compassion-based prosocial behavior more strongly than norm-based prosocial behavior. The effectiveness of the compassion training was further supported by an increase in positive mood and compassionate feelings and a decrease in negative mood. Future research with longer training and bigger sample sizes needs to ascertain how long lasting these effects are and who is benefitting from compassion training. Clinical research for example suggests that some people find compassion-focused imagery distressing (Gilbert, 2007; Rockliff, Gilbert, McEwan, Lightman, & Glover, 2008) and thus do not benefit from it. Furthermore it needs to be investigated whether long-term compassion training leads to stronger increases in specific types of prosocial behavior and whether this effect can also be observed in everyday life behavior. As the interpersonal effects were directed towards total strangers and transferred to situations outside the training context, compassion training could have great societal impact when implemented in institutions of daily life.

Materials and Methods

Participants. In Experiment 1, that aimed to validate the newly developed ZPG, we investigated 68 healthy female volunteers (aged 18–35 years; mean = 25.18; years of education after the 16th birthday: 2– 15 years; mean = 6.54). In Experiment 2, that aimed to assess the effect of a short-term compassion training workshop on prosocial behavior as measured using the newly developed ZPG, we investigated 69 healthy female volunteers (age: 18–34 years; mean: 23.69). Only female participants were included in Experiments 1 and 2 because of better performance in emotional tasks (Kring & Gordon, 1998) and higher self-reported empathy in women (Baron-Cohen & Wheelwright, 2004). All participants came from the University of Zurich and the surrounding community and were recruited through local advertisement and internet postings. The advertisements for Experiment 2 asked for people interested in mental training but never mentioned the word compassion. All participants completed the Toronto Alexithymia Scale (TAS; Bagby, Parker, & Taylor, 1994), the Beck's Depression Inventory (BDI; Beck, Steer, Ball, & Ranieri, 1996) and socio-demographic questions online. Only when they met the following inclusion criteria they were contacted via telephone: aged 18–35 years, TAS < 60, BDI < 18, right hander and no contraindication for fMRI. Importantly, possible participants of Experiment 2 were additionally not allowed to have prior experience with mental compassion training or the method of loci. On the phone, participants were given information about the timing but, importantly, in case of Experiment 2, not about the specific content of the study and underwent a structured psychological interview (screening questions for axis-I disorders and psychotic disorders of the Structured Clinical Interview for DSM Disorders (SCID; German version: Wittchen, Zaudig, & Fydrich, 1997). Woman with current psychiatric illnesses were excluded from the study. For Experiment 2, allocation to the compassion-training and memory-training (control) group depended on slot availability and time of the participants. 35 participants entered the compassion-training group and 34 participants entered the memory-training group. 28 participants from the compassion group and 32 participants from the memory group completed the study. One participant of the compassion group was eliminated from the analysis as data on the ZPG was missing. The majority of the dropout in the compassion group (5/7) occurred before the training. Furthermore, the seven participants that dropped out of the study did not differ in age, $t(32) = 0.57$, $p > .05$, years of education, $t(32) = 1.75$, $p > .05$, empathic concern, $t(32) = -0.81$, $p > .05$, alexithymic symptoms, $t(32) = 1.70$, $p > .05$, depressive symptoms, $t(32) = -0.75$, $p > .05$, prosocialness, $t(32) = -0.79$, $p > .05$, compassionate feelings, $t(32) = -0.99$, $p > .05$, and general positive, $t(32) = 1.42$, $p > .05$, and negative affect, $t(32) =$ -

0.06, $p > .05$, from the participants that finished the study, thus excluding selective dropout in the compassion group.

Table 2.2

	Validation sample (N = 68)	**Compassion training sample (n = 27)**	**Memory training sample (n = 32)**
Age	25.18 (4.08)	24.74 (4.22)	22.66 (3.86)
Highest completed education	Apprenticeship: 5 (7.5%) High school: 34 (50.8%) University: 28 (41.8%)	Apprenticeship: 3 (11.1%) High school: 18 (66.7%) University: 6 (22.2%)	Secondary school: 1 (3.1%) Apprenticeship: 1 (6.3%) High school: 24 (75%) University: 4 (12.6%) PhD: 1 (3.1%)
Education (years after 16th birthday)	6.54 (2.87)	5.48 (2.44)	5.06 (2.54)
Prosocialness[1]	64.03	60.75	64.50
Empathic concern[2]	27.64	27.08	27.19
Alexithymia[3]	41.24	39.41	41.16
Depression[4]	6.13	6.04	4.53

*Note***.** Sample characteristics [1]Prosocialness Scale (Caprara, Steca, Zelli, & Capanna, 2005; range: 16-80) [2]Empathic Concern Subscale Interpersonal Reactivity Index (IRI; Davis, 1983; range: 7-35) [3]Toronto Alexithymia Scale (TAS; Bagby et al., 1994). > 60 clinically relevant) [4]Beck's Depression Inventory (BDI; Beck et al., 1996). > 18 clinically relevant)

The compassion group and the memory group did not differ in age, $t(57) = 1.98$, $p > .05$, years of education, $t(57) = 0.64$, $p > .05$, prosocialness, $t(54) = 1.91$, $p > .05$, empathic concern, $t(54) = .1$, $p > .05$, alexithymic symptoms, $t(57) = 0.89$, $p > .05$, or depressive symptoms, $t(57) = 0.87$, $p > .05$. There was also no difference in the distribution of type of education between the samples ($X^2 = 3.06$, $p > .05$). See Table 2.2 for sample characteristics. The study was approved

by the Research Ethics Committee of Zurich (‘‘Kantonale Ethikkommission des Kantons Zürich – Spezialisierte Unterkommission Psychiatrie, Neurologie, Neurochirurgie’’; E- 25/2008) and was performed according to the Declaration of Helsinki. All participants gave written informed consent after having received a full description of the study.

Measures

Zurich Prosocial Game. A novel game, the Zurich Prosocial Game (ZPG), was developed that allows for repeated assessment of prosocial behavior and for parsing the influence of reciprocity, cost, and distress on prosocial behavior. The participants’ task is to navigate a virtual character through a maze and reach a treasure in a limited amount of time. Each treasure is worth 0.50 Swiss francs ($ 0.50). At the same time, participants see the virtual character of an ostensible co-player from another research institute in Europe who is also trying to reach a treasure. Importantly, the two players do not share the same paths in the maze and do not compete for the same treasure. Thus, in principle, the game can be played while completely ignoring the other player. Participants are told that in each round of the game they are connected via the internet with a new co-player who is sitting in a different research institute in Europe. At the onset of each round, the participant and the ostensible co-player select one of two paths. While the players move their virtual character through the maze, red and blue gates fall on the paths that can block the participant and the co-player. Each of the two players is equipped with red and blue keys with which they can open the corresponding gates. When the co-player runs out of keys, participants can use their own keys to open the gates for them. Importantly, participants cannot delay their help to observe the progression of the game (i.e., whether they need their keys themselves) as the virtual characters become inactive before the next gate falls and thus cannot reach the treasure anymore. During each trial, participants can see how many gates are still going to fall, which and how many keys they and the co-player still possess, and how much time is left (for a screenshot of the game display, see Figure 2.3). When playing the game, participants wear headphones as sounds convey distress cues in the distress trials and add emphasis to events on the screen (e.g., sound when a gate is falling). Importantly, to reduce demand effects, participants are never told that the purpose of the game is to help the co-player. Instead, the instructed aim of this computer game is to reach a goal, the treasure, in a short amount of time to optimize monetary winnings.

Different trial types were introduced to probe the effect of different factors on prosocial behavior. First, to assess the influence of reciprocity on prosocial behavior, no-reciprocity and reciprocity trials were created. In the no-reciprocity trials, participants had the opportunity to help the co-player while knowing that the co-player would not have any opportunity to reciprocate as either no gates at all or no gates that the participant could not open with her own keys were still going to fall. In the reciprocity trials, participants had the opportunity to help the co-player after the co-player had helped them earlier in the trial. In these trials, participants could also see that there would not be any opportunity for the co-player to reciprocate. By designing the trials this way and by changing the co-player for each trial, we excluded the possibility of participants helping because they anticipated that they might need the co- player to reciprocate later on.

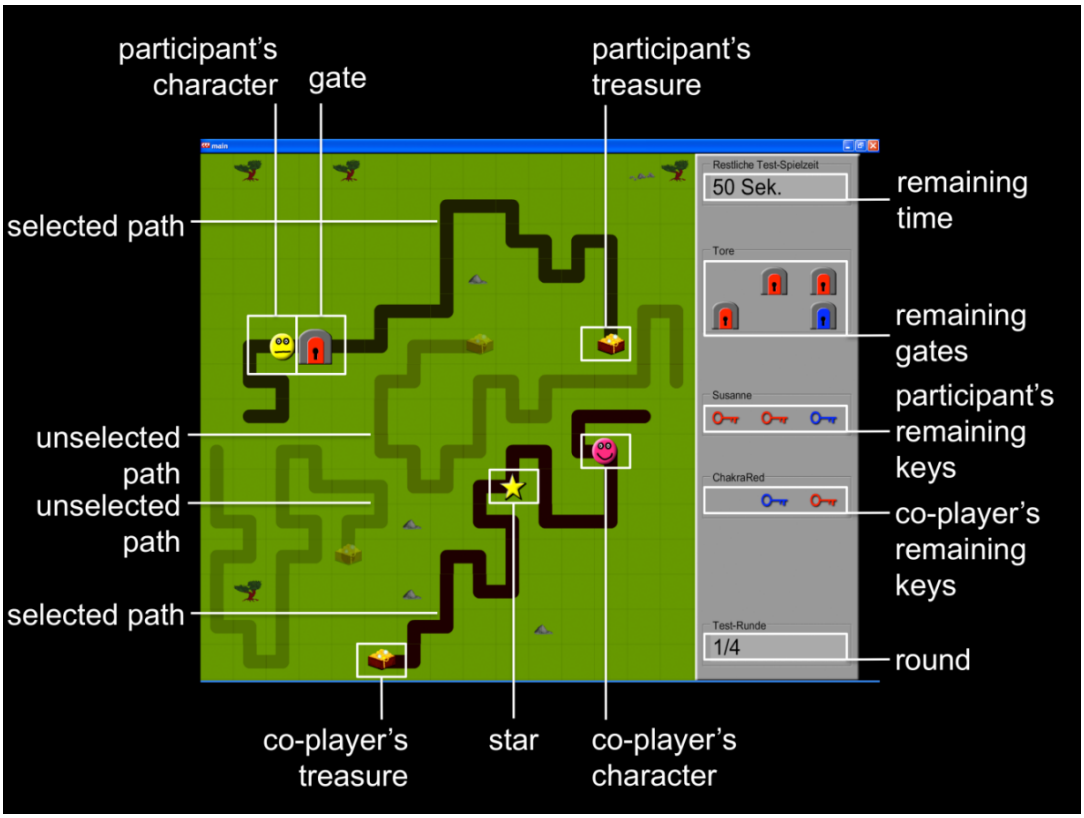


Figure 2.3. Labeled screenshot of the ZPG. Participants move their virtual character forward by clicking with the mouse on the field in front of it. Usage of keys in order to open the blocking gates occurs by mouse click on the key matching the gate's color. Collection of stars also occurs by clicking on them with the mouse.

To assess the influence of helping cost on prosocial behavior, there was a low- and high-cost variant of all trial types. In the high- cost variant, participants knew that after they helped the co-player, there would be a 25% chance that they would need the donated key to reach the treasure themselves; in the low-cost variant, players knew that they could donate keys without risking needing them later themselves, the only cost in this condition being loss of time.

Finally, to investigate the effect of distress cues, when the co- player's virtual character was blocked, it either a) started to cry and sweat, as implemented by visual changes in the virtual character and by crying sounds that participants heard over headphones (distress cues) or b) gave no distress cues. This resulted in a 2 x 2 x 2 factorial design with the three factors reciprocity (no reciprocity, reciprocity), cost (low, high) and distress (no distress, distress).

One game consisted of nine trials, one of each type, plus one trial in which no helping was necessary to reduce the affordance of the game. Trial types appeared in random order with the restriction that the first reciprocity trial could appear at the earliest as the third trial. This restriction was introduced to reduce the likelihood of an anchoring effect (helping agreed upon behavior) being introduced by experiencing a helpful co-player right away. At the beginning of the game, participants were first given written and verbal instructions and asked five questions probing their comprehension of the game. Then participants completed four practice trials to familiarize themselves with the handling of the game and to determine individual reaction time thresholds. To offset individual differences in speed and proficiency with computer games, the individual time limit for all trials of a given game was set at the average time the individual required to reach the treasure in the four practice trials plus 5 s. Furthermore, to control for the possibility that inter-individual differences in helping might be due to differences in participants' allocation of attention to their own and the co-player's path, we let a star appear randomly on some trials. The star yielded 0.20 Swiss francs ($ 0.20) when picked up. This was expected to result in the allocation of attention to the whole display, as the star could appear anywhere. If participants collected stars but refrained from helping, attentional influences on helping behavior could most certainly be ruled out.

Risk perception control. As inter-individual differences in the behavior in the high cost trials could be brought on by differences in risk preferences, i.e., participants differ in their perception of the risk of not reaching the treasure in the high cost trials, we assessed risk preferences. First, we asked participants on an eight-point scale how risk-seeking they are and, second, we

presented them with seven lotteries where the amount that can be won (6 Swiss Francs) stays the same but the amount that can be lost varies (1–7 Swiss Francs). Participants can decide for each lottery whether they want to play it or not. The computer then randomly picks one lottery and the outcome of this lottery is paid out to the participants if they had decided to play it. The number of lotteries accepted is an index for risk preferences.

Engagement with the game. Participants were asked after playing the game to indicate on a five-point scale how engaged they were when playing the game. A high engagement of the participants would indicate that they were emerged in the game and diminish the probability that demand effects and strategic decision-making influenced prosocial behavior in the ZPG.

Dictator game. To assess the convergent validity of the ZPG, participants played the dictator game (Hoffman et al., 1994). Based on our reasoning about norm-based and compassion-based prosocial behavior in the introduction, we did not expect an exceedingly high correlation between the ZPG and the dictator game, but still, as both tasks assess variants of prosocial behavior, a sufficiently high correlation to maintain that the ZPG indeed measures prosocial behavior. Participants were again told that they would be paired with another player from another research institute in Europe. In the dictator game, based on random assignment, participants are endowed with 80, 120, or 160 points that they can split between themselves and an ostensible co-player who has no points. Points are later converted to money with a conversion scheme of one point equaling six, four, or three Swiss rappen (or ‘‘Swiss penny’’), respectively.

Memory task. To assess the divergent validity of the ZPG, we used a memory task that was later used as an outcome measure for the memory training group in the intervention experiment. Participants were presented with 34 words on the computer screen and were asked to memorize them and their sequence. Each word appeared for four seconds and words were separated with a 2 s presentation of a crosshair. After the presentation, a word file opened and participants had five minutes to remember as many words as possible in the correct sequence.

Effectiveness measures – Experiment 2. To measure the effectiveness of the compassion-training workshop, we assessed the difference in mood and compassionate feelings reported before and after compassion training. Participants completed the Positive and Negative Affective Scale (PANAS; Watson et al., 1988) and the Compassionate Love Scale (CLS; Sprecher & Fehr, 2005). If effective, compassion training should lead to increases in positive mood, compassionate feelings, and possibly to a decrease in negative mood.

Experiment 1

Procedure. All participants gave written informed consent after having received a description of the study. They were told that they would play interactive computer games via the internet with other participants in different research institutes across Europe in order to investigate cross-cultural differences in interpersonal behaviors. In reality, there were no co-players; the ostensible co-players' behavior was pre-programmed. Participants were seated in front of a computer and the experimenter provided oral and written instructions to the Zurich Prosocial Game (see below). Participants then answered five questions testing their comprehension of the game. The experimenter then checked the answers to ensure that participants fully understood the rules of the game. Then participants put on headphones and the experimenter ostensibly logged the participant into the game network. An abbreviated version of the instructions appeared on the computer screen and participants were asked to enter a freely chosen nickname to play the game. When ready, participants started the game and played four practice rounds after which they again had the opportunity to ask questions before playing the actual game. After finishing the Zurich Prosocial Game, the experimenter provided oral and written instructions explaining the dictator game (see below) and ostensibly logged the participant into the game network again. After playing the dictator game, participants completed a memory task (see below), filled out questionnaires (see below) and completed a lottery task to assess risk preferences (see below). Participants also completed a task in the magnetic resonance (MR) scanner and other unrelated tasks (results to be reported elsewhere). All participants were debriefed after the study was completed.

Experiment 2

Procedure. Participants came to the lab one to two weeks prior to the training for their pre-training measurement (pre-test) and two to five days after the training workshop for their post-training measurement (post-test). The pre- and post-training measurements were identical, except that risk-preferences were only assessed at pre-test. A detailed description of the measurement procedure with respect to the ZPG can be found in the documentation of Experiment 1. Briefly, participants first played the ZPG and the dictator game, both under the assumption that they were playing the games with other participants in research institutes across Europe. Afterwards, participants completed the memory task and a lottery task to assess risk preferences (only at pre-test). Then, in contrast to the validation sample in Experiment 1, at

post-test participants filled out questionnaires that were to probe the effectiveness of the compassion training (see below). Participants also completed a task in the MR scanner and other unrelated tasks (results to be reported elsewhere). Thus, the newly developed Zurich Prosocial Game was assessed in the context of several other non-helping tasks, which further helped to reduce possible demand effects of the compassion training. Participants were asked to continue practicing after the training (see below for training details) in the days before post-test. To facilitate continuation, we offered a one-hour guided evening training session on each of these days. Participants were debriefed after the end of the study.

Compassion and memory training. The compassion group attended a one-day training to learn a compassion- enhancing technique developed in Buddhist contemplative traditions. This compassion meditation technique (called ‘‘Metta’’ in Pali) aims to foster an attitude of loving kindness, emotional positivity, benevolence, and friendliness towards oneself and others (Carson et al., 2005; Salzberg, 1995). An experienced meditation teacher with over ten years of teaching experience led the training workshop. The training involves sitting in an upright position and developing warm, positive feelings sequentially towards oneself, a beloved person, a neutral person, a person one has difficulties with, and all human beings by imagining each while silently repeating sentences like ‘‘May you be happy’’ or ‘‘May you be safe’’ and cultivating these positive emotional attitudes towards the visualized persons. The training day was held in silence and lasted for six hours in which mental training was sometimes done while sitting and sometimes while walking. The mental training periods were usually between 15 and 30 minutes long. There was a 45 minutes lunch break in between. During the course of the training, the target of the compassion meditation changed in the following succession: oneself, beloved person, neutral person, difficult person, all human beings. Ultimately, this should lead to an attitude of emotional positivity, benevolence, and friendliness towards oneself and others (Salzberg, 1995). Thus, as in compassion-focused therapy (Gilbert, 2010), compassion here is trained as a skill. In contrast to a momentary induction of empathic concern through instruction to feel for a specific person in distress in a specific situation, compassion training aims at permanently changing one’s motivation and attitude towards others in general.

The memory control group underwent a one-day training workshop in the method of loci, a technique used to memorize items in an ordered sequence (Bower, 1970; Verhaeghen & Marcoen, 1996). An experienced memory technique teacher with over ten years of teaching experience led the training workshop. The method of loci involves linking a series of locations (e.g., a learned route through Zurich) with a series of specific items (a list of words) by creating

visual mental images that combine each item with a location. For example, in order to remember the word ‘‘egg,’’ one would imagine a big fried egg hanging down from the towers of the cathedral in Zurich. During recall, one recreates the images by mentally walking from one location to the next. This particular mnemonic technique was chosen for the control group as it contains most elements also needed in the compassion training workshops: People need to actively engage in inner mental processing and to create active mental images and specific associations between items. The difference is that the memory group focuses purely on improving cognitive rather than affective skills. Participants of both groups were asked to continue training in the days before post-test (1–3 days) and keep a diary of their practice. They were asked to join the daily offered one-hour evening training sessions or, if this was not feasible, to train at home.

Acknowledgments

We thank Alexandra Freund for helpful comments and suggestions on previous versions of this manuscript, Astrid Hopfensitz for valuable ideas during task development, and George Margaris for programming the ZPG.

Author Contributions

Conceived and designed the experiments: TS SL OK. Performed the experiments: SL OK. Analyzed the data: SL. Wrote the paper: SL TS OK.

3. Functional Neural Plasticity and associated Changes in Positive Affect after Compassion Training

Corresponding publication:

Klimecki, O., Leiberg, S., Lamm, C., Singer, T. (2012). Functional neural plasticity and associated changes in positive affect after compassion training. *Cerebral Cortex*. doi: 10.1093/cercor/bhs142. By permission of Oxford University Press.

Abstract

The development of social emotions such as compassion is crucial for successful social interactions as well as for the maintenance of mental and physical health, especially when confronted with distressing life events. Yet, the neural mechanisms supporting the training of these emotions are poorly understood. To study affective plasticity in healthy adults, we measured functional neural and subjective responses to witnessing the distress of others in a newly developed task (Socio-affective Video Task, SoVT). Participants' initial empathic responses to the task were accompanied by negative affect and activations in the anterior insula and anterior medial cingulate cortex – a core neural network underlying empathy for pain. Whereas participants reacted with negative affect before training, compassion training increased positive affective experiences, even in response to witnessing others in distress. On the neural level, we observed that, compared to a memory control group, compassion training elicited activity in a neural network including the medial orbitofrontal cortex, putamen, pallidum, and ventral tegmental area – brain regions previously associated with positive affect and affiliation. Taken together, these findings suggest that the deliberate cultivation of compassion offers a new coping strategy that fosters positive affect even when confronted with the distress of others.

Compassion can be defined as “the emotion one experiences when feeling concern for another’s suffering and desiring to enhance that individual’s welfare” (Keltner & Goetz, 2007). In other words, compassion is composed of an other-related affective component and a prosocial motivational component. In the context of the present study, we investigated compassion as rooted in a contemplative tradition aimed at cultivating feelings of warmth, love, and care towards a sequence of people in a state of quiet concentration (Fredrickson et al., 2008; Leiberg, Klimecki, & Singer, 2011). More specifically, the training focused on the practice of loving kindness (‘Metta’ in Pali) that aims to foster an attitude of emotional positivity, benevolence and friendliness, as it is naturally experienced towards close loved persons like one’s own mother or children (similar to strengthening modes of affiliation). During the training, this attitude was extended towards all persons independent of context and preferences.

In order to reliably investigate how cultivating feelings of warmth and prosocial motivation influences neural responses when exposed to the suffering of others, we developed a new paradigm – the Socio-affective Video Task (SoVT) – in which participants saw short video sequences based on real-life situations depicting others in distress. Experiencing feelings of loving kindness towards another person’s affliction is best described with the term ‘compassion’. In other words, compassion can be regarded as applying loving kindness (as trained) in response to the suffering of others (implemented by the SoVT). Thus, although the concepts of compassion and loving kindness are separable as the object they refer to can differ, we will for reasons of simplicity and clarity generally refer to compassion and compassion training in the present paper.

Even though closely related, compassion (feeling *for* someone) differs from empathy, which is defined as feeling *with* someone, i.e. sharing the other person’s emotion (Batson, 2009; deVignemont & Singer, 2006; Singer & Lamm, 2009). As such, empathic responses to witnessing another person’s suffering usually involve shared negative affective experiences. Indeed, most neuroscientific studies on empathy to date have focused on the domain of vicarious pain. These studies revealed that a network including the anterior insula (AI) and the anterior medial cingulate cortex (aMCC) is involved both in the first-hand experience and the vicarious experience of pain (Fan, Duncan, de Greck, & Northoff, 2011; Lamm, Decety, & Singer, 2011). Furthermore, activity in aMCC has been repeatedly shown to be modulated by the experience of unpleasantness (Rainville, 2002). Despite these advances in empathy research, the neural correlates underlying compassionate feelings are still poorly understood.

Furthermore, it has remained unresolved which neural processes are relevant for strengthening such prosocial emotions.

Although neural plasticity has been studied for decades (Pascual-Leone, Amedi, Fregni, & Merabet, 2005), the majority of research has favored the investigation of neural changes induced by the training of motor and cognitive abilities (Karni et al., 1995; Olesen, Westerberg, & Klingberg, 2004), while the study of neural plasticity underpinning socio-affective competences has long been neglected. To date, research in the domain of affective plasticity has focused mainly on short-term changes associated with fear conditioning and fear extinction (Gottfried & Dolan, 2004; Schiller, Levy, Niv, LeDoux, & Phelps, 2008). The ability to cope with difficult situations is crucial for successfully maneuvering through our social lives, as well as for maintaining mental and physical well-being (Gross & Muñoz, 1995). Therefore, the lack of studies focusing on the neural mechanisms underlying the plasticity of positive social affect is surprising. The neuroscience of compassion is actually still in its infancy and has, so far, only been studied cross-sectionally (Beauregard, Courtemanche, Paquette, & St-Pierre, 2009; Immordino-Yang, McColl, Damasio, & Damasio, 2009; Kim et al., 2009; Lutz, Brefczynski-Lewis, Johnstone, & Davidson, 2008; Simon-Thomas et al., 2011). To fill this gap, we investigated functional neural plasticity induced by compassion training in healthy adults in a series of functional magnetic resonance imaging (fMRI) experiments which combine multiple approaches with the aim of obtaining measures which are robust in terms of replicability, validity, and generalizability (Figure 3.1 A and Supplementary Figure S3.2).

First, we assessed the neural networks underlying the deliberate generation of compassion in an expert long-term meditation practitioner immersed in different types of compassionate states. Subsequently, we carried out two compassion training experiments in meditation-naïve participants (total N = 46) from which we derived independent regions of interest (ROIs). These findings were used to qualify the neural changes observed in a third longitudinal experiment which compared compassion training in 28 female participants with an active control group (n = 30) who received training in a memory-enhancing technique (Bower, 1970). To reliably assess participants´ compassionate responses, we specifically developed the Socio-affective Video Task (SoVT), which measures empathy, positive affect and negative affect in response to witnessing videos depicting people in distressing situations (high emotion videos, HE) or showing people performing everyday activities (low emotion videos, LE). The task was composed of three parallel video sets which can be used interchangeably to avoid habituation induced through repeated measurement, thus making it optimal for longitudinal designs.

In keeping with previous findings on the neural substrates of empathy, we expected that empathy for HE videos would be accompanied by activations in the AI and the aMCC, which are core-regions of the empathy for pain network (Fan et al., 2011; Lamm et al., 2011). With regard to the effects of compassion training, we had the following hypotheses: Given that other-related feelings of warmth and concern are key components of compassion, we anticipated that loving kindness training would increase activity in brain regions previously associated with love (Bartels & Zeki, 2000; Aron et al., 2005) or affiliation (Vrticka, Andersson, Grandjean, Sander, & Vuilleumier, 2008; Strathearn, et al., 2009), such as the (medial) orbitofrontal cortex (mOFC), the putamen, the middle insula and the ventral tegmental area/substantia nigra (VTA/SN, which are mentioned together as the spatial resolution of the present fMRI data does not allow for an unambiguous distinction between both midbrain structures). We furthermore expected these neural changes to be paralleled by an increase in positive affect and a decrease in negative affect.

Materials and Methods

Subjects. We investigated a total of 94 right-handed healthy female volunteers between 18 and 35 years of age (mean age = 24.3, SD = 4.17) (Figure 3.1 A, see Main Experiment, Participants in Supplementary Material). The study was approved by the Research Ethics Committee of Zurich (''Kantonale Ethikkommission des Kantons Zürich – Spezialisierte Unterkommission Psychiatrie, Neurologie, Neurochirurgie''; E-25/2008) and was carried out in compliance with the Declaration of Helsinki. All subjects gave written informed consent, were paid for their participation and were debriefed after the study was completed.

Measures. A new measure of social affect, the SoVT was developed to assess social emotions in response to short videos of 10-18 s. To avoid habituation and recognition effects, participants were tested with a different randomly selected parallel video set at each measurement time (for details, see "Task Development" in Supplementary Material). This assured that all video sets (A, B, and C) were equally viewed at all measurement time points by all groups of participants. The videos were presented without sound and depicted raw material cast for news or excerpts from documentaries. Each of the video sets was composed of 12 high emotion (HE) videos which depicted persons in distress and 12 low emotion (LE) videos which served as a high level baseline and showed people performing everyday activities (Figure 3.1 B). Videos were shown

in blocks of 3 HE or 3 LE videos (resulting in a total of 24 videos per measurement point). Following each video, participants rated on three different scales how much empathy, positive affect, and negative affect they experienced when watching the videos (scale: from zero to ten, presented in a pseudo-randomized order for 6 s each). Null events displaying a fixation cross were presented for 10 s after each mini-block. Participants were instructed to observe the videos and to pay attention to their emotional response. At post-training, participants were encouraged to make use of their trained competences when viewing the videos.

The memory task consisted of three parallel lists of 34 words, which enabled us to test participants with a different list at each measurement time point. Words were presented sequentially on a computer screen for 4 s each, followed by a 2 s fixation cross. Subsequently, subjects were given 5 min to fill in a table with the recollected words, if possible, in the correct sequence. To obtain questionnaire measures of affective traits, we measured empathy by means of the Interpersonal Reactivity Index (Davis, 1983) and compassion by means of the Compassionate Love Scale (Sprecher & Fehr, 2005).

Procedure. Pre-training measurements were collected one to two weeks prior to training and post-training measurements were collected two to twelve days after training. Both measurements consisted of behavioral and fMRI components and were essentially identical, apart from specific questionnaires collected only at pre- or post-training in the context of a related study (Leiberg et al., 2011). In the behavioral part of the study, participants played the Zurich Prosocial Game (Leiberg et al., 2011), performed the memory task, and completed a series of questionnaires. Prior to being scanned with the SoVT, participants received written and verbal instructions on the task and performed three practice trials. The total duration of the video task in the fMRI scanner was 15 min, consisting of two 7.5 min sessions separated by an anatomical scan, resulting in a total scan time of 20 min. To validate the SoVT, 33 participants were shown video set A, 31 viewed video set B, and 30 saw video set C at pre-training.

Compassion Training. The compassion training group attended a one-day course of loving kindness meditation (6 h) led by a teacher with more than 10 years of teaching experience. The course was aimed at fostering benevolent and friendly attitudes towards oneself and other persons based on techniques from Eastern contemplative traditions (Salzberg, 2002). Ultimately, the goal was to develop compassion as a generalized prosocial feeling and motivation, independent of particular persons or situations. The entire course was carried out in a quiet contemplative state adopted in an upright seated posture or during silent walking. Participants received a theoretical introduction, followed by guided practices (15-30 min) of strengthening

feelings of warmth and care through the visualization of a close loved person (e.g., a child). Participants sequentially extended these feelings towards themselves, a close person, a neutral person, a person in difficulty, and finally towards strangers and human beings in general. At the end of the course, participants practiced modulating the intensity of compassion towards a loved person. After the day of training, participants filled out a debriefing questionnaire which asked several quantitative questions and gathered qualitative reports about the participants' training experiences (see Supplementary Tables S3.1 and S3.2). These additional reports confirmed that, on average, participants liked the loving kindness training and could successfully engage in the different practices taught during that day. In between pre- and post-measurement, participants practiced the trained method during evening classes, during three laboratory measurements (45 min each) in the context of another study, and at home. Participants were asked to record their number of home training hours. Twenty-six out of 28 participants in the compassion group returned their questionnaires to us indicating a mean of 5.7 practice hours (range 45 min to 11.17 hours). We cannot be certain about the reliability of these data, as we did not use electronic measurement tools to objectively assess home practice. Furthermore, participants were encouraged to practice throughout their daily lives and, indeed, several reported engaging in the trained method in everyday life situations (e.g., using public transport). Finally, the number of training hours did not correlate with any of the dependent variables. Taken together, we know certainly that in addition to the one intense day of compassion training (6h) and the informal training at home and in every-day life, participants practiced a minimum of two additional supervised training hours.

Memory Training. Like the compassion-training group, the memory control group also attended a one-day course (6 h) held by a teacher with more than ten years of experience in teaching the Method of Loci, a technique to memorize items in an ordered sequence (Bower, 1970). After initial practice with a self-chosen sequence of ten locations, participants learned a fixed sequence of 34 locations (a route through Zurich). Lists of items were remembered by creating vivid mental images that associated each item with a location. Item lists were recalled by mentally revisiting each location on the route and by bringing to mind the previously created mental images. Thus, both training groups engaged in active mental imagery: the memory group trained a purely cognitive ability and the compassion group trained an emotion related to attachment and affiliation. The post training debriefing questionnaire revealed that participants of the memory training group also enjoyed the mnemonic training and were also able to apply the taught memory technique (see Supplementary Table S3.1). As with the compassion group,

the memory group also participated in evening classes or trained at home. Since the return rate of the self-reported training hours of the memory control group was not reliable (only 22 out of 30 persons provided this information), this data was not analyzed.

Behavioral Data Analysis. Self-reports of empathy, positive affect, and negative affect in response to the SoVT, and performance in the memory task were analyzed using repeated measures multivariate analyses of variance (MANOVAs; with Pillai's trace statistics), followed-up by analyses of variance (ANOVAs) and *t*-tests. All statistical tests were two-tailed and used an alpha level of 0.05. Correlation coefficients were computed using Pearson's *r*. Correlations were compared using Dunn and Clark's *z* test. At post-training measurement, in 4 participants ratings of one of two runs were lost due to a program error reducing subjective ratings from the initial 24 videos to 12 videos for these 4 participants. Since the recall data from the memory task were missing from one subject in the compassion group and from one subject in the memory group, the total sample size for the memory task was n = 56.

fMRI Data Acquisition and Analyses. General linear models and parametric modulations of ratings were estimated on the first level for each participant after standard preprocessing (realignment, normalization and smoothing). The resulting contrasts were then entered into second level random effects ANOVAs and *t*-tests (for details, see "Main Experiment, fMRI Data Acquisition and Analyses" in Supplementary Material). Neuroimaging results are reported at $p < .05$, corrected for family wise errors (FWE) using Gaussian random fields theory and a cluster extent threshold of k = 5 voxels. When restricting the search volume to a priori defined regions of interest (ROIs), results are reported with a FWE corrected significance threshold of $p < 0.05$ using small volume correction (SVC). For completeness, we also report brain regions activated at $p < 0.001$, uncorrected in the tables. Since these activations were not in the a priori defined regions of interest and did not survive whole-brain correction, we do not draw conclusions about them in the present paper.

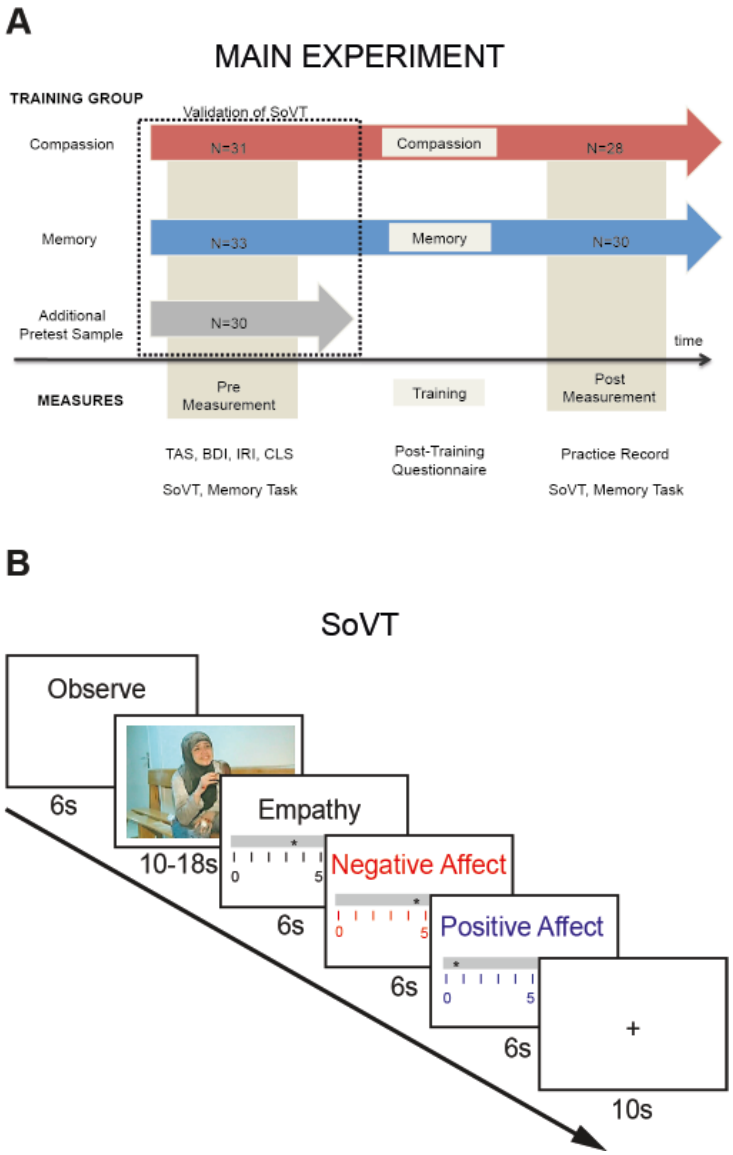


Figure 3.1. Experimental design. *(A)* Main Experiment. At pre-training measurement, a total of 94 women were scanned by means of fMRI while viewing one of three parallel video sets from the newly developed Socio-affective Video Task (SoVT). 30 of 33 participants completed the memory training, 28 of 31 participants completed the compassion training and 30 participants were included as an additional validation sample at pre-training measurement. Pre-training measurement included the trait questionnaires Toronto Alexithymia Scale (TAS, Bagby et al., 1994), Beck's Depression Inventory (BDI, Beck et al., 1996), Interpersonal Reactivity Index (IRI, Davis, 1983) and Compassionate Love Scale (CLS, Sprecher & Fehr, 2005), as well as the memory task and fMRI measurements in response to the SoVT. After the training day, self-report data about the training were collected in form of a post-training questionnaire. At post-training measurement participants handed in their practice records, performed the memory task and were scanned while viewing the SoVT. *(B)* SoVT. In the main experiment, each fMRI

measurement presented participants with 12 high emotion (HE) and 12 low emotion (LE) videos (duration 10-18 s). After each video, participants rated their subjective experience of empathy, positive affect and negative affect. Every mini-block of three HE or LE videos was followed by a fixation cross (displayed for 10 s).

Results

Validation of the Socio-affective Video Task (SoVT)

Analyses focusing on the validity of the SoVT confirmed the following (for details, see Task Development in Supplementary Material, Supplementary Figure S3.2 and Supplementary Tables S3.3 - S3.5): Watching HE compared to LE videos evoked more negative and less positive feelings. Furthermore, participants reported feeling more empathy towards other people in distressing situations compared to everyday situations (Figure 3.2 A-C). Pearson correlations between self-report ratings on the SoVT and questionnaire measures of empathy (assessed by the Interpersonal Reactivity Index, IRI, Davis, 1983) and compassion (measured by the Compassionate Love Scale; CLS, Sprecher & Fehr, 2005) confirmed the external validity of the SoVT. Empathy ratings for HE videos were significantly correlated with total empathy scores ($p < .05$, $r = .23$): This effect was driven by the subscales empathic concern ($p < .01$, $r = .34$) and fantasy ($p < .001$, $r = .42$). Compassionate love was correlated with empathy for LE and HE videos ($p < .01$, $r = .29$, and $p < .001$, $r = .36$, respectively), as well as with positive affect for LE ($p < .01$, $r = .29$) and HE videos ($p < .05$, $r = .26$).

Neuroimaging results (reported at $p < .05$, FWE corrected) revealed that HE videos compared to LE videos, as well as parametric modulations of negative affect and empathy, induced activations in a network comprising areas in the occipital, temporal and parietal lobes, as well as in subcortical areas including the amygdala, thereby replicating previous findings on neural activity typically involved in the processing of negative pictures or videos (Goldin, McRae, Ramel, & Gross, 2008; Vrticka et al., 2011). Conversely, LE compared to HE videos elicited stronger activation in the posterior insula, the mOFC, the fusiform and the precentral gyrus. In contrast to negative affect ratings, positive affect ratings mapped to the mOFC, a brain region consistently implicated in the experience of hedonic feelings as well as in the encoding of value and reward (Kringelbach & Berridge, 2009; Rangel & Hare, 2010; Grabenhorst & Rolls, 2011). Finally, the validity of the SoVT as a measure of empathy for suffering in others could be

confirmed by showing that neural activity parametrically modulated by empathy ratings in response to HE videos mapped to activations in the AI (x = 34, y = 26, z = 2; Z = 4.14, p < .05, SVC), and the aMCC (x = 0, y = 10, z = 34; Z = 4.66, p < .05, SVC) – both core regions involved in empathy for others' suffering (Lamm et al., 2011; Fan et al., 2011) (Figure 3.2 D and Table S3.5).

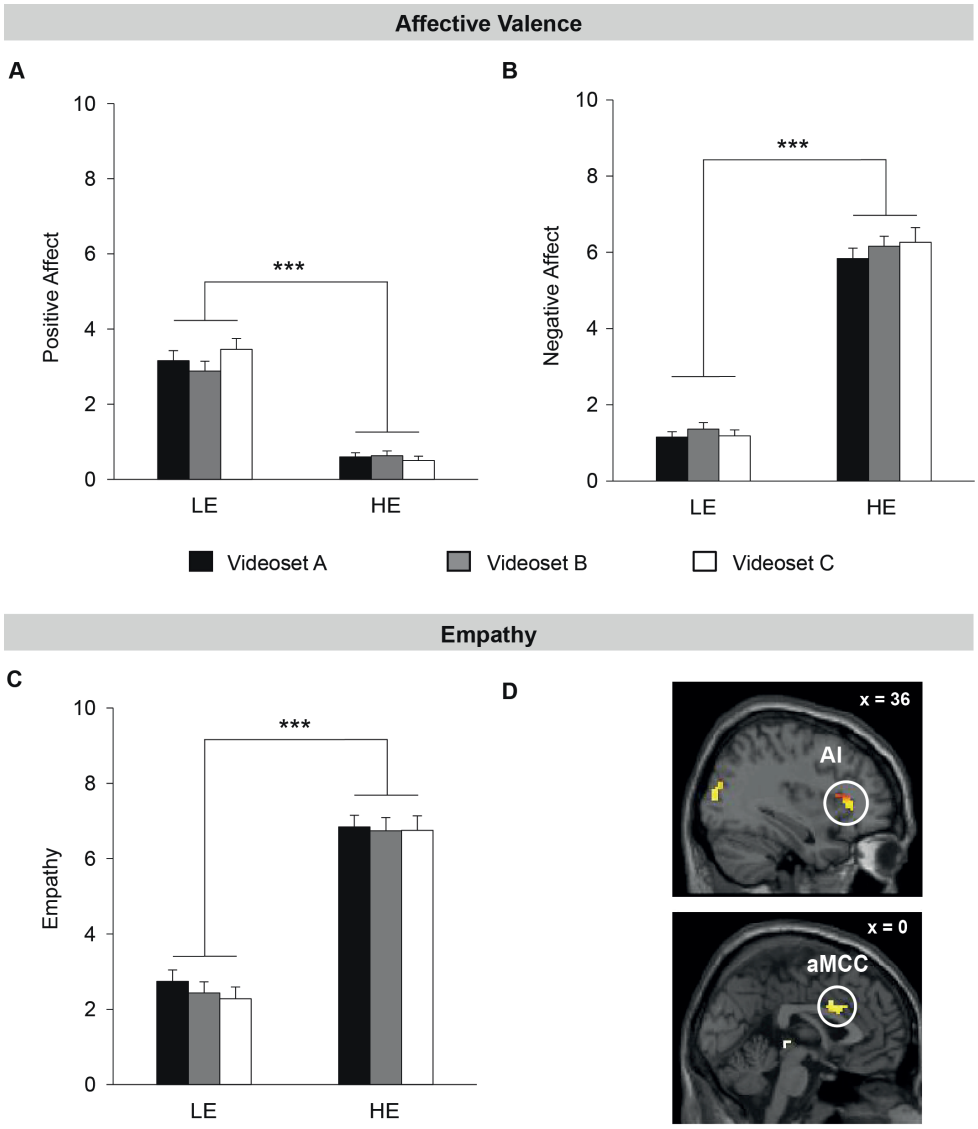


Figure 3.2. Subjective emotional experiences and related neural activations evoked by the SoVT. *(A-C)* Average subjective ratings of positive affect, negative affect and empathy in response to LE and HE video stimuli at pre-training (zero, no affect; ten, maximum affect). Whereas the video sets (A, B and C) did not differ on any of the ratings, the effect of video type (LE and HE videos) was highly significant for each dependent variable. Error bars indicate standard error of mean (SEM); ***, p < .001 *(D)* Small volume correction (SVC) with a functional activation map from a recent meta-analysis on empathy for pain (Lamm

et al., 2011) revealed that activations in the AI (x = 34, y = 26, z = 2; Z = 4.14, p < .05, SVC) and the aMCC (x = 0, y = 10, z = 34; Z = 4.66, p < .05, SVC) were parametrically modulated by subjectively experienced empathy for HE videos during the SoVT. Color-coded activations with brighter colors indicating lower p values were rendered on a Montreal Neurological Institute (MNI) template in neurological orientation. Inset x values indicate stereotactic coordinate of the shown slice in MNI space. AI, anterior insula; aMCC, anterior medial cingulate cortex.

Training-induced Plasticity

In order to assess whether short-term compassion training would lead to increased positive and decreased negative affect during vicarious experiences, we performed a 2 x 2 x 2 repeated measures MANOVA with video type (HE and LE videos) and time (pre- and post-training) as within subject factors and training group (compassion and memory) as a between subject factor. The dependent variables were empathy, positive affect and negative affect ratings. The main effects of video type ($F(3, 54)$ = 320.70, p < .001, η^2 = .95) and time ($F(3, 54)$ = 2.80, p < .05, η^2 = .14) were significant, and the main effect of training group showed a trend towards significance ($F(3, 54)$ = 2.70, p = .06, η^2 = .13). Significant interactions were found for video type x time ($F(3, 54)$ = 4.70, p < .01, η^2 = .21), video type x group ($F(3, 54)$ = 3.27, p < .05, η^2 = .15) and time x group ($F(3, 54)$ = 5.64, p < .01, η^2 = .24); the triple interaction video type x time x group was not significant ($F(3, 54)$ = 2.11, p = .11). As expected, the difference between HE and LE videos was confirmed for all three dependent variables in paired t-tests (all $t(57)$ ≥ 15.18, all p < .001). Univariate ANOVAs showed that the main effect of time was significant for empathy ($F(1, 56)$ = 4.57, p < .05) and positive affect ($F(1, 56)$ = 4.36, p < .05). Furthermore, all three interactions were significant for the dependent variable empathy (all $F(1, 56)$ ≥ 8.15, all p < .01). Positive affect only showed a time x group interaction ($F(1, 56)$ = 8.66, p < .01), whereas contrary to our expectations, no training effects were found for negative affect. Follow-up paired t-tests revealed that in the compassion group, empathy ratings for LE videos were higher at post- than at pre-training ($t(27)$ = 3.71, p < .01) and independent t-tests showed that post-training empathy ratings for LE videos were higher in the compassion than the memory group ($t(56)$ = 4.18, p < .001). Similarly, positive affect ratings were higher at post- than at pre-training in the compassion group ($t(27)$ = 2.68, p < .05) and independent t-tests determined that positive affect ratings at post-training were higher in the compassion than the memory group ($t(42.26)$ = 2.72, p < .05). All other effects were not significant. In summary, compassion training (but not memory training) led to significant increases in ratings of positive affect for LE and HE videos

and to an increase in empathy ratings for LE videos (Figure 3.3 A-C). Of note, self-reports of 26 out of 28 participants from the compassion group in a post-scan debriefing indicated that when viewing the videos after the training, participants seemed to have adopted a compassionate stance towards HE videos (mean = 5.38, SD = 0.88) and LE videos (mean = 4.26; SD = 1.38, on a scale from 1, minimum to 7, maximum). This suggests that these participants indeed engaged in compassion when viewing the SoVT at post-training.

Interestingly, in the compassion group the correlation between empathy and negative affect at pre-training measurement was reduced at post-training, which may be accounted for by the increase in empathy ratings in response to LE videos (pre-training: $r = .69$, $p < .001$; post-training: $r = .15$, $p = .44$; Dunn and Clark's z test $Z = 2.96$, $p < .01$). No such effect was observed for the memory group (pre-training: $r = .65$, $p < .001$; post-training: $r = .82$, $p < .001$; $Z = 1.04$, $p = .3$), and no other correlations showed significant changes. Furthermore, the effectiveness of the memory control training was confirmed (see Analysis of the Memory Task in Supplementary Material Text and Figure 3.3 D-E).

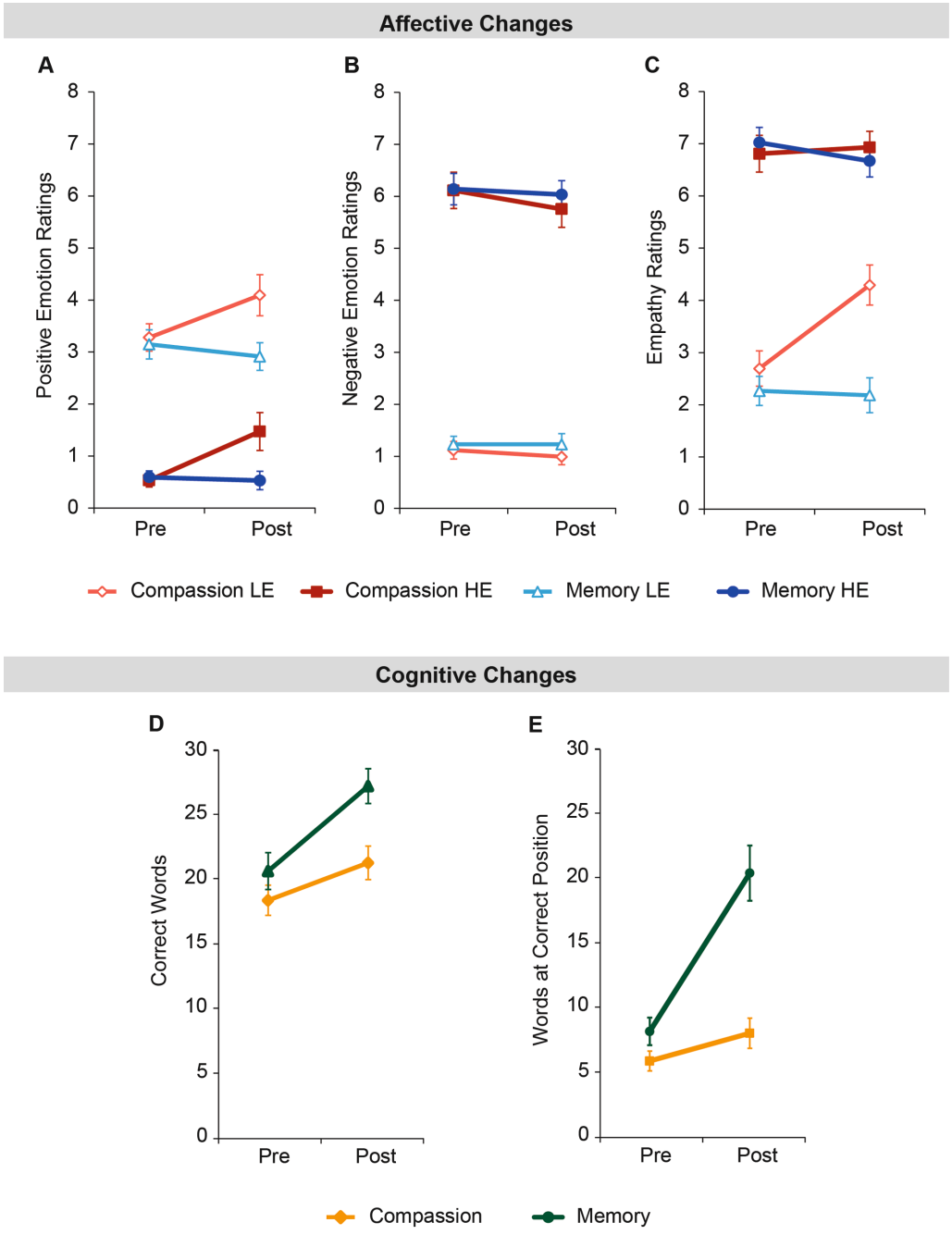


Figure 3.3. Behavioral effects of compassion training and memory training. *(A)* Self-reported positive affect in response to LE and HE videos increased after compassion training, but not after memory training. *(B)* No significant changes were observed for negative affect. *(C)* Compassion, but not memory training, increased empathy towards people in LE videos. *(D)* Memory and compassion training improved the number of correctly remembered words. Importantly, at post-training measurement, the memory group remembered significantly more correct words than the compassion group. *(E)* Memory training induced an increase in the number of words remembered in the correct position. Error bars indicate SEM.

As the field of compassion research is just emerging, we performed multiple experiments to assure replicability, validity, and generalizability of the training based neuroimaging results. Prior to conducting the main experiment, we performed three studies to determine the core neural signatures underlying the generation of compassionate feelings. In particular, these independent experiments included different populations (novice meditators and an expert practitioner) across different countries (the Netherlands and Switzerland) using different task parameters (immersion in compassionate states without an external task and brain responses to the SoVT). Through such a multi-study and multi-method approach, we aimed to reduce measurement noise, secure replicability and validity, and optimize the generalizability of the findings. First, to obtain external validity, we tested a long-term expert practitioner (more than 35 years of daily practice in compassion enhancing mental training techniques) while he immersed himself in three different compassionate states during 15 fMRI sessions acquired over three days (see Expert Data in Supplementary Material). We observed that across all three compassion-enhancing techniques, the generation of strong, as compared to weak, feelings of compassion consistently increased activations in subcortical structures including the caudate nucleus and striatum, and cortical regions, including the mOFC and supplementary motor areas ($p < .05$, FWE corrected; Supplementary Table S3.6). In a next step, we derived independent ROIs based on two previously conducted short-term compassion training experiments in meditation-naïve participants. Specifically, the first experiment focused on the neural correlates reflecting subjective intensity ratings of newly trained compassionate states, while the other experiment measured neural changes related to compassion training in responses to others' suffering in the newly developed SoVT (total N = 46, see ROI construction in Supplementary Material and Supplementary Table S3.7). Interestingly, the neural effects of short-term compassion training were consistent with activations in the expert practitioner in the right mOFC, VTA/SN, pallidum, and putamen (Figure 3.4 orange boxes), as well as bilaterally in the precentral gyrus, amygdala, and thalamus and on the left side in the pallidum, putamen, middle occipital gyrus, supramarginal and superior temporal gyrus.

Finally, we analyzed functional neural plasticity induced by compassion training compared to memory training within the independently derived ROIs (Supplementary Table S3.8). To this end, we computed first-level contrast images of post > pre-training activation separately for LE and HE videos (ΔLE and ΔHE videos) and entered these into a 2 x 2 repeated measures random effects ANOVA with the within subject factor video type (ΔLE and ΔHE videos) and the between subject factor group (compassion and memory training). We first tested the effect of

group across both ΔLE and ΔHE videos. This revealed that the compassion group had stronger activations in the right putamen (x = 28, y = -6, z = 6; Z = 3.08, p < .05, SVC) and pallidum (x = 26, y = -8, z = 4; Z = 2.88, p < .05, SVC). The interaction [(Compassion ΔHE - ΔLE) > (Memory ΔHE - ΔLE)] revealed activations in the right mOFC (x = 12, y = 38, z = -18; Z = 2.73, p < .05, SVC). As the main focus of this study was to determine whether compassion training was able to alter brain responses elicited when exposed to the distress of other people, we also directly contrasted training-related changes when watching HE videos between the compassion and the memory group. This revealed that compared to memory training, compassion training was associated with activations in the mOFC, VTA/SN, putamen, and pallidum (all right lateralized, p < .05, SVC; Figure 3.4). Furthermore, training-related changes were not significant in response to LE videos, although Figure 3.4 suggests that there was a trend in the same direction as for HE videos.

Interestingly, the memory group showed increased neural activity in response to LE videos in the right hippocampus (x = 40, y = -16, z = -22; Z = 2.57, p < .05, SVC), a region relevant for memory processes (Squire et al., 2004) and the right fusiform gyrus (x = 40, y = -16, z = -24; Z = 2.71, p < .05, SVC). No other training-based activation changes survived correction in the compassion-related ROIs.

Parametric analyses on the changes in ratings revealed that the overall increase in positive ratings (across HE and LE videos) in the compassion group correlated with activity changes in the right putamen (x = 34, y = -2, z = -2; Z = 2.99, p < .05, SVC). Changes in negative affect ratings correlated with activations in the left supramarginal gyrus (x = -50, y = -26, z = 20; Z = 2.8, p < .05, SVC), the left superior and the left middle temporal cortex (x = -44, y = -44, z = 10; Z = 2.99 and x = -46, y = -46, z = 8; Z = 3.14, respectively, both p < .05, SVC), while no correlations were observed with changes in empathy ratings.

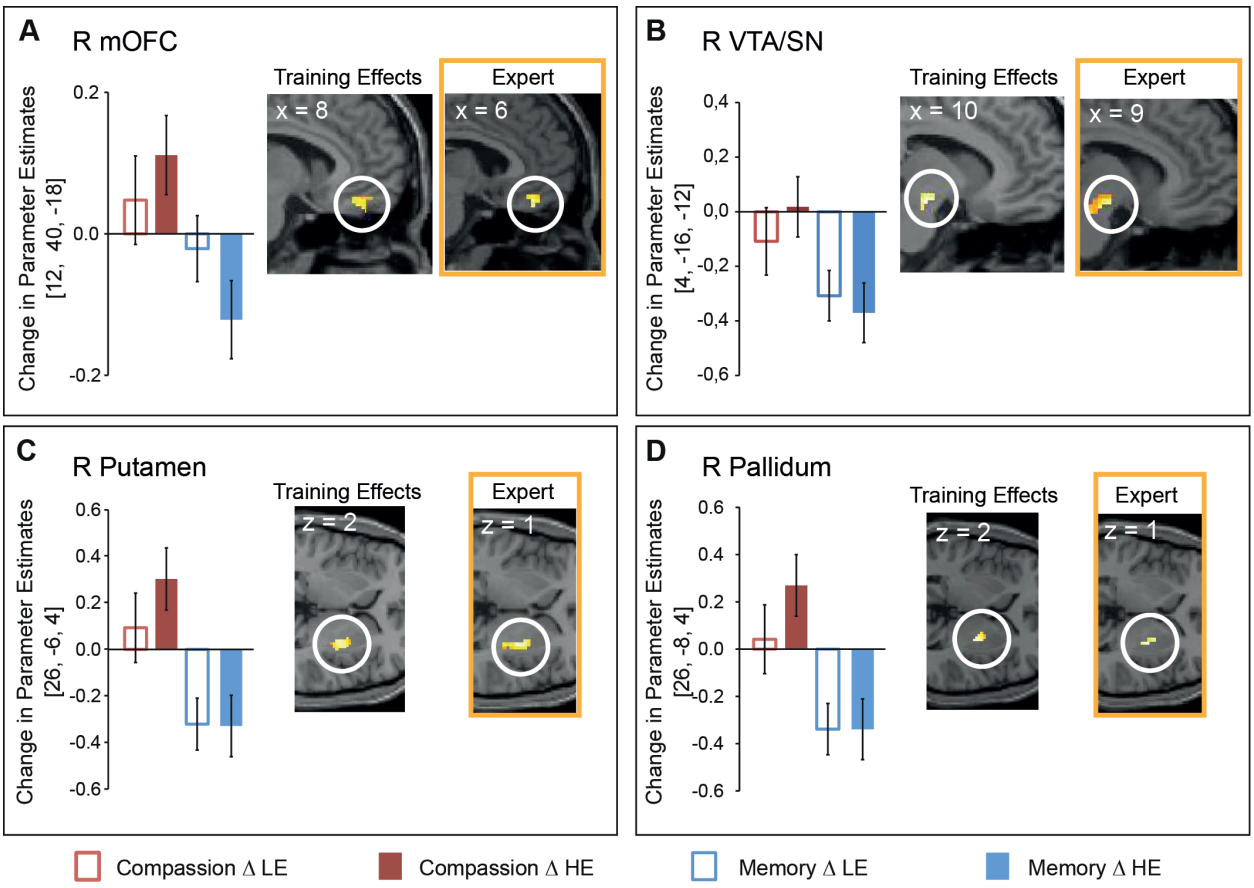


Figure 3.4. Effects of compassion training and memory training on neural responses to HE videos. The contrast [Compassion Δ HE > Memory Δ HE] revealed activations (p < .05; SVC) in *(A)* the right medial orbitofrontal cortex (mOFC), *(B)* the right ventral tegmental area/ substantia nigra (VTA/SN), *(C)* the right pallidum and *(D)* the right putamen. Bar charts show the change in parameter estimates in the depicted independent region of interest; error bars denote SEM Orange boxes show neural activations (p < .05; FWE corrected) of an expert practitioner immersed in three compassionate states [high > low degree] across 15 sessions depicted within independent regions of interest. Color-coded activations with brighter colors indicating lower p values were rendered on an MNI template in neurological orientation. Inset x/z values indicate stereotactic coordinate of the shown slice in MNI space.

Discussion

The present study investigated functional neural plasticity induced by compassion training. In particular, we studied which subjective and neural changes in response to distressed others were induced by compassion or memory training. As fMRI compatible paradigms for the repeated probing of affective responses to social situations are rare, we developed and validated a new fMRI task, the SoVT. The SoVT was optimized for repeated measurements: We generated three video sets which could be used interchangeably. The employed videos depicted short everyday life scenes of people in low or high emotional states. Self-reports collected after each video allowed the dissociation of empathy, positive and negative affect in response to other people's suffering.

Essentially, we observed that compassion training compared to memory training a) increased positive affect, even in response to others' suffering, and b) was associated with stronger activations in a specific neural network including the mOFC, the pallidum, the putamen, and the VTA/SN – brain regions previously implicated in positive valuation (Kringelbach and Berridge, 2009), as well as love (Bartels & Zeki, 2000; Bartels & Zeki, 2004; Aron et al., 2005; Beauregard et al., 2009) and affiliation (Vrticka et al., 2008; Strathearn et al., 2009). These results are particularly striking, since participants were exposed to aversive (HE) videos, which at pre-training actually elicited negative affect and empathy-related activations in the AI and the aMCC – core regions of the empathy for pain network (Lamm et al., 2011; Fan et al., 2011) that have typically been modulated by subjective experience of unpleasantness (Rainville, 2002). Finally, the parallel nature of the neural pattern in several experiments including novices and an expert, lends further support for the replicability and validity of the observed compassion related brain activation pattern.

In light of previously reported findings, the observed activation pattern underlying the training of compassion accords with neural correlates of love, affiliation and positive affect. More precisely, romantic (Bartels & Zeki, 2000; Aron et al., 2005) and maternal love (Bartels and Zeki, 2004) have been shown to induce stronger activations in the VTA/SN, the middle insula, the putamen and the pallidum. Similarly, activations in the striatum and the VTA were observed in response to smiling faces and this activation was reduced in individuals with avoidant attachment (Vrticka et al., 2008). In addition, secure attachment in mothers seeing photographs of their happy children has been related to activity in the ventral striatum and the mOFC (Strathearn et al., 2009). More generally, mOFC activations were consistently associated with positive affect, such

as reward value and pleasure (Kringelbach & Berridge, 2009; Rangel & Hare, 2010; Grabenhorst & Rolls, 2011), which ties in well with the pre-training observation of a correlation between ratings of positive affect and mOFC activity. Furthermore, the present results are in line with cross-sectional studies focusing on compassion either through instruction-induced effects in naïve participants (Beauregard et al., 2009; Kim et al., 2009) or through testing expert meditation practitioners (Lutz et al., 2008). For example, the involvement of VTA/SN and mOFC was also observed when participants were instructed to adopt a compassionate attitude towards sad faces (Kim et al., 2009) and activations in the VTA, the pallidum and the middle insula were observed when adopting an attitude of love towards disabled persons (Beauregard et al., 2009). Middle insula activation was reported to be crucial when expert practitioners immersed themselves in compassion meditation while hearing human cries of distress (Lutz et al., 2008). Note that in the present study, middle insula activation was also observed to be parametrically modulated by changes in positive affect ratings in the compassion group, albeit at a lower threshold of $p = .001$. The present results extend the previous cross-sectional findings in an important way as they provide, for the first time, longitudinal evidence that activations in the mOFC, the VTA/SN, the pallidum, and the putamen (extending into middle insula) are stronger in response to distressing video stimuli after compassion training but not following memory training.

Neural changes were particularly pronounced for HE videos, although the parameter estimates in Figure 3.4 show similar but weaker changes in response to LE videos. On the behavioral level, this effect was paralleled by increased self-report ratings of positive affect in response to both HE and LE videos. The overall increase in positive affect through the cultivation of warmth and benevolence in the present study adds to the finding that daily self-reports of positive emotions were augmented after loving kindness training (Fredrickson et al., 2008). Extending previous results, the present study actually shows that positive emotions were not only increased in response to everyday life situations, but even in response to witnessing the distress of others. The observed increase in reported positive affect – even when exposed to the suffering of others – suggests that persons trained in compassion can encounter social situations in general and distressing situations in particular with positive, other-oriented affect. A review of the qualitative self-report data after the training day of the compassion group gave us some more information about the qualia associated with the generation of loving kindness and compassion (see Supplementary Table S3.2). For instance, participants indicated that they liked, *"the feeling of warmth"; "the wonderful, fulfilling feeling to wish others well"; "feelings of love,*

security, shelter"; "to think of other people and […] give them kindness"; "a feeling of happiness that arises." They further reported, *"I felt an enormous heat. It was nice to feel that one could create something like that inside oneself."; "It was easier than I thought to generate feelings of benevolence towards a stranger."* Although these qualitative self-reports cannot be quantified and used in analyses, they nonetheless give us valuable indications that loving kindness training strengthened feelings of warmth and care towards oneself and others. Finally, the indication that participants indeed used the practice of loving kindness when exposed to the SoVT at post-training measurement, suggest that participants re-interpret the meaning of empathy: instead of experiencing empathy as a distressing and aversive emotion that may ultimately result in withdrawal tendencies (Batson et al., 1983; Eisenberg et al., 1989), compassion training may have induced an understanding of empathy as a rather positive feeling associated to empathic concern (Batson, 2009) and warmth.

Finally, contrary to our expectation, compassion training only increased positive affect, but did not decrease negative affect. In line with previous reports of co-activated positive and negative affect (e.g., Larsen, McGraw, & Cacioppo, 2001), our finding underlines the need to measure positive and negative affect independently (see also Watson et al., 1988; Cacioppo & Berntson, 1994). Indeed, the co-activation of positive and negative affect may be a crucial prerequisite for successful coping with stressors (Larsen, Hemenover, Norris, & Cacioppo, 2003). Future studies may help to dissociate the neural substrates underlying positive and negative affective experiences associated with social and vicarious emotions.

In a broader context, the current results extend our understanding of existing strategies known to be useful for coping with negative affect such as extinction learning (Gottfried & Dolan, 2004) or different forms of explicit emotion-regulation strategies such as cognitive reappraisal (Ochsner, Bunge, Gross, & Gabrieli, 2002; Kalisch, 2009), emotion suppression (Levesque et al., 2003; Phan et al., 2005) or detachment (Walter et al., 2009). In line with previous reports on beneficial effects of compassion training on mood (Fredrickson et al., 2008) and health (Pace et al., 2009), the current data show that responding to potentially distressing situations with compassion benefits the person who experiences it through strengthening positive affect. Furthermore, compassion benefits others through fostering prosocial behavior (Leiberg et al., 2011). By consequence the generation of compassion may not only help others but also be beneficial for oneself by providing a potentially powerful coping strategy. However, to situate compassion in relation to the existing framework of emotion regulation strategies, future studies

are needed which directly compare different emotion regulation techniques with compassion generation.

Regarding the duration of training, the current study ranged at the lower end of interventions, therefore examining short-term functional neural plasticity. Future studies are needed to map out the behavioral, subjective, and neuronal changes that occur after longer compassion training interventions of several months to years. In addition, it would be interesting to understand the role specific neurotransmitter systems play in strengthening compassion. Although the spatial resolution used in the present experiments did not allow us to disentangle the VTA and the SN, the relatively stronger activation for compassion compared to memory training in this area is interesting, since the VTA and the SN are rich in dopamine (Fields, Hjelmstad, Margolis, & Nicola, 2007) and the SN is rich in oxytocin binding sites (Loup, Tribollet, Dubois-Dauphin, Pizzolato, & Dreifuss, 1989). In addition, the VTA is a key region mediating the rewarding effects of opiates (Burgdorf & Panksepp, 2006). Future neuropharmacological intervention research and Positron Emission Tomography (PET) with specific ligands might help to unravel the contribution and interaction of dopaminergic reward (Schultz, 2010), oxytocin-related affiliative processes (Strathearn et al., 2009), and opiate-mediated well-being (Burgdorf & Panksepp, 2006) in the generation of compassion.

In conclusion, the present findings provide evidence for functional neural plasticity induced by compassion training and thus constitute a crucial step in the endeavor of understanding neural plasticity in the affective domain. The results suggest that compassion training can serve as a new, powerful method for enhancing positive affect in response to adverse situations. This is in accordance with findings of beneficial effects of compassion training for strengthening personal resources (Fredrickson et al., 2008) and health (Pace et al., 2009). As positive affect has been found to be associated with resilience (Fredrickson, Tugade, Waugh, & Larkin, 2003), future studies should explicitly address whether compassion training can serve as a buffer against stress, depression, and anxiety, and which physiological substrates mediate this effect. Importantly, in comparison to other techniques for increasing positive affect, compassion benefits both, the person who experiences it (through strengthening positive affect) and the recipient of compassion (through fostering prosocial motivation). In the long-run, this promises to have major implications for the development of interventions for normal and clinical populations.

Funding

This work was supported by the University of Zurich ('Neural Foundations of Empathy Learning', Forschungskredit, grant number 53990601 to O.M.K.), grants to T.S. from the Neuroscience Center Zurich, the Swiss National Science Foundation (SNF NCCR Affective Sciences: 'Neuronal and developmental mechanisms of human empathy'; SNF NCCR Neural Plasticity and Repair: 'Plasticity of the empathic and attentional brain'), the Foundation for Research in Science and Humanities at the University of Zurich ('Neural foundations of empathy training'), and a Francisco J. Varela Research Award to S.L. from the Mind and Life Institute. The research leading to these results has also received funding from the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013/ ERC Grant agreement number 205557 to T.S.).

Acknowledgments

We are deeply grateful to Dr. Matthieu Ricard for all the fruitful discussions that resulted in this research. Furthermore, we gratefully acknowledge Dr. Rainer Goebel and Dr. Bettina Sorger for their contribution in Maastricht. Finally, our gratitude goes to the teachers Catherine Felder, Annette Rentz-Lühning, Jotika Hermsen and Ariya Ñani, as well as to Fred von Allmen.

Supplementary Material

Task Development

Subjects, Measures, Analysis and Results. Stimulus material for the three parallel video sets of the Socio-affective Video Task (SoVT) was validated based on a normative study with 265 healthy volunteers (age M = 22.57 years, SD = 4.86; 133 male) who each rated 40 to 51 videos (from a total of 342 videos). Two participants of initially 267 participants had to be excluded since they did not adhere to the task instructions. Videos were selected from documentaries and unedited material cast for news and consisted of either high emotion (HE) videos depicting distressed people, or low emotion (LE) videos showing people performing everyday activities. Subjects were seated in a multi-computer laboratory and rated each video (presented via z-Tree (Fischbacher, 2007) without sound) on valence, arousal, empathy and ease of assessment on a nine point scale ranging from zero to eight. In addition, the distractor items fear, anger, and disgust were rated. Based on videos rated as 'easy to assess' (above midpoint of the scale), we constructed three parallel video sets (A, B or C) with 12 HE and 12 LE videos, each. To confirm the matching of the three video sets, we set up 3 x 2 multilevel linear models for each dependent variable separately (valence, arousal and empathy) with the factors video set (3 levels: video set A, B or C) and video type (2 levels: HE or LE videos). The mixed between and within subject design of the study was taken into account by modeling the variable 'subject' as a correlated random effect. Confirming the parallel nature of the video sets, no differences were found between the video sets (all $F \leq 0.83$, all $p \geq .44$). Furthermore, as expected, the differences between HE and LE videos were confirmed for all three dependent variables (all $F \geq 643.12$, all $p < .001$). Item properties are depicted in Supplementary Table S3.3.

Expert Data

Subject. The expert practitioner (65 years of age) had over 35 years of extensive experience and daily practice with different Tibetan meditation techniques (amounting to more than 50 000 hours of training), including the three compassion states investigated in the current study. He had been taught by masters in Nepal, and acquired ample knowledge about the philosophical background of contemplative practices by living as a monk in a Nepalese monastery. The study was approved by the local Ethics Committee of the Faculty of Psychology and Neuroscience at Maastricht University and was carried out in compliance with the Declaration of Helsinki. The practitioner gave written informed consent.

Procedure. The expert practitioner came to the laboratory three days in a row. On these three days, he completed 15 fMRI sessions in which he was asked to immerse himself in three different compassionate states (unconditional compassion, loving kindness, pain compassion; see below). Each session consisted of nine meditation blocks of 60s each, interspersed with ten resting baseline blocks of 30s each. The nine meditation blocks differed in the intensity (low, medium, or high) of the specific compassionate state in which the expert practitioner was asked to immerse himself. Four different random sequences of the meditation blocks were used with the restriction that no more than two blocks of the same intensity could occur in a row. On Day One, the expert practitioner completed one session each of unconditional compassion and pain compassion, as well as five sessions of loving kindness. On Day Two, he completed three sessions of pain compassion and one session of loving kindness. On Day Three, he completed three sessions of unconditional compassion and one session of loving kindness.

Compassionate States. The expert practitioner immersed himself in three different compassion states: The state of unconditional compassion is characterized by a very pronounced experience of unconditional love, compassion and benevolence towards all living beings in conjunction with a strong motivation to act for the benefit of others. It is a non-referential state that does not focus on a particular person or object. The state of loving kindness (the state in which the participants of the present study were trained in) also includes powerful experiences of unconditional love and benevolence and a strong wish that people may be safe, healthy and happy. However, here the feeling concentrates on particular beings. The expert practitioner imagined members of his family, children and animals and let his love flow towards them. Compassion focused on suffering is characterized by similar feelings as loving kindness and also involves imagining particular beings. However, here the focus lies on the suffering of others.

It involves imagining the suffering of particular beings, feeling it very strongly, and developing empathic concern and wishing for the relief of that suffering.

fMRI Data Acquisition and Analyses. MRI data from the expert practitioner were collected at the Maastricht Brain Imaging Centre (Maastricht University) using a 3 T Siemens Allegra MRI scanner equipped with a standard quadrature birdcage head coil. T2*-weighted images (34 slices) were acquired axially using an echo planar imaging (EPI) pulse sequence (TR = 2s, TE = 30ms, flip angle = 90°, FoV = 224 × 224 mm^2; matrix size = 64 × 64 x 64mm^3, voxel size = 3.5 × 3.5 × 3.5 mm^3, gap = 0.5 mm). The first two volumes were discarded to allow for signal equilibration.

SPM8 software (Wellcome Trust Centre for Neuroimaging, London) on Matlab 7.10 (MathWorks, Natick, MA) was used for analyzing the fMRI data of the expert practitioner. Preprocessing included spatial realignment, normalization of the expert's functional data to the Montreal Neurological Institute (MNI) EPI template, and smoothing with a Gaussian kernel of 6 mm full-width at half-maximum (FWHM). Then, a general linear model that accounted for temporal autocorrelations (AR(1)) and low frequency drifts (with a cut-off period of 256s adapted to the long block length) was applied to the time series of each voxel. This model included four regressors for each of the 15 sessions (7 sessions loving kindness, 4 sessions unconditional compassion, 4 sessions compassion for suffering): low immersion in compassionate state, medium immersion in compassionate state, high immersion in compassionate state, resting baseline. Realignment parameters were entered into the design matrix to further correct for residual effects of head motion. Regressors were convolved with a hemodynamic response function. To investigate the effect of immersion in the compassionate state, we computed the contrast [high immersion in compassionate states > low immersion in compassionate states], abbreviated as [high > low]. The results are reported at $p < .05$, FWE-corrected for the whole brain using Gaussian random fields theory and a cluster extent threshold of k = 10 voxels.

ROI construction

<u>fMRI Experiment A</u>

Subjects. In Experiment A, 25 participants were initially enrolled, but three dropped out, leaving a final sample of 22 healthy women from the University of Maastricht and the surrounding community (age M = 25.4 years; SD = 3.5) who fulfilled the inclusion criteria described in the

main methods section. All participants were right-handed, native speakers of Dutch or German, and had no prior experience with meditation. The study was approved by the local Ethics Committee of the Faculty of Psychology and Neuroscience at Maastricht University and was carried out in compliance with the Declaration of Helsinki. All participants gave written informed consent and were paid for their participation.

Measures and Procedure. All participants took part in a one-day course on loving kindness meditation led by an experienced meditation teacher (> 10 years of teaching experience). In between the training and the testing, participants practiced at home or during two measurements in the lab (in the context of another study). Two to eight days after the course, participants were scanned while they internally generated compassionate feelings in six blocks of 46s each. Compassion blocks were interspersed with seven baseline blocks in which participants had to count backwards for 30s. After each compassion block, participants rated how well they had been able to generate the feeling of compassion on a four-point scale. Participants completed two high level compassion sessions and three intensity modulation sessions (see Expert Procedure). Since the training focused on loving kindness, participants immersed themselves in this particular state.

fMRI Data Acquisition. MRI data were collected at the Maastricht Brain Imaging Centre (Maastricht University) using a 3 T Siemens Allegra MRI scanner equipped with a standard quadrature birdcage head coil. T2*-weighted images were acquired axially using an echo planar imaging (EPI) pulse sequence (TR = 2s, TE = 30ms, flip angle = 90°, FoV = 224 × 224 mm^2; matrix size = 64 × 64 x 64mm^3, voxel size = 3.5 × 3.5 × 3.5 mm^3, gap = 0.5 mm). The first two volumes were discarded to allow for signal equilibration.

fMRI Data Analysis. FMRI data analysis was performed with Statistical Parametric Mapping Software (SPM8; Wellcome Department of Imaging Neuroscience, London, UK) in Matlab 7.10 (The Mathworks, Natick, MA). Preprocessing included spatial realignment, spatial normalization to the EPI template and spatial smoothing with an isotropic Gaussian kernel (FWHM 6 mm). A general linear model that accounted for low-frequency drifts (with a cut-off period of 256s adapted to the long block duration) and first-order temporal autocorrelations was computed for each participant. The model included three regressors for each session: (1) compassion meditation, (2) backwards counting and (3) the parametric regressor of block-by-block compassion ratings. Realignment parameters were included to account for subject movement. To test for brain activation related to subjective experiences of compassion, we computed first-

level contrasts representing parametric modulations of BOLD signal by compassion ratings. The resulting contrast-images were smoothed with an isotropic Gaussian kernel (FWHM 8 mm).

fMRI Experiment B

Subjects. In Experiment B, the final sample consisted of 24 right-handed healthy female volunteers (age M = 24.58 years, SD = 4.24) who fulfilled the inclusion criteria described in the main methods section. Of the 25 participants who completed the whole training study, one had to be excluded due to excessive head movement (more than 3 mm).The study was approved by the Research Ethics Committee of Zurich (''Kantonale Ethikkommission des Kantons Zürich – Spezialisierte Unterkommission Psychiatrie, Neurologie, Neurochirurgie''; E-25/2008) and was carried out in compliance with the Declaration of Helsinki. All participants gave informed consent and were paid for their participation.

Measures, Procedure and Compassion Training. The measures and procedures were mostly identical to those employed in the main experiment with the main exception that participants received a minimum of two days (12h) of guided compassion training. For the analysis, we compared neural responses to HE videos after 6 h of compassion instruction (Time 1) with neural responses to HE videos after participants had trained in compassion for another whole day (Time 2).

fMRI Acquisition and Analyses. Acquisition and preprocessing of fMRI data was identical to the procedures described in the main methods section. To test for compassion-induced changes in response to the SoVT, we modeled the three regressors LE videos, HE videos and rating phase on the first level for both time points and sessions and determined the contrast [compassion-induced change in response to HE videos for Time 2 >Time 1]. First-level contrasts were smoothed with 8 mm FWHM before being entered into second-level analysis. In order to derive the regions of interest based on Experiment A and Experiment B, we set up a second-level random effects model with the between subject factor "Experiment" that contained the contrasts [subjective experiences of compassion] (Experiment A) and [compassion-induced change in response to HE videos for Time 2 > Time 1] (Experiment B). Regressors for both conditions were set to one. Using marsbar (Brett et al., 2002), neural activation was extracted at $p < .005$, uncorrected and subsequently restricted to MNI regions (Brett et al., 2002) (without cerebellum, as we had no hypothesis about activations in this region) and the VTA/SN (Filbey et al., 2007). Regions with a cluster extent > 810 mm^3 (corresponding to 30 voxels) were used for subsequent analyses (Supplementary Table S3.7).

Main Experiment

Participants. Only female participants were recruited due to previously reported gender differences in social emotions on the behavioral and neural level (Baron-Cohen and Wheelwright, 2004; Singer et al., 2006, repectively). Participants had to meet the following inclusion criteria: Toronto Alexithymia Scale (Bagby et al., 1994) < 60, Beck's Depression Inventory (Beck et al., 1996) < 18, no contraindication for fMRI, and no current psychiatric illness as determined by a Structured Clinical Interview for DSM Disorders (SCID) (Wittchen et al., 1997) via phone. Allocation of participants to training groups was performed without revealing the specific content of the study and depended on the participant's temporal availability. Five participants were excluded due to unreturned pre-training questionnaire data or due to technical problems during the pre-training fMRI measurement. The study was completed by 28 of 31 participants in the compassion group, 30 of 33 participants in the memory group and 30 participants in an additional training group, whose longitudinal results will be reported elsewhere. All participants had no prior experience in the training they received. Neither the compassion and memory group, nor the three different groups viewing video set A, B, and C differed on the socio-demographic and questionnaire variables at pre-training measurement (Supplementary Tables S3.9 and S3.10).

fMRI Data Acquisition and Analyses. A 3T Philips Achieva scanner with an eight channel sense head coil was used for data acquisition. Blood oxygenation level dependent (BOLD) contrast functional images were acquired using a T2*-weighted single-shot echo-planar imaging (EPI) pulse sequence (TR = 1.5s, TE = 30ms, flip angle = 90°, FoV = 240 x 228 mm^2, slice gap = 1 mm). We obtained 29 transverse slices (voxel size = 3 x 3 x 3 mm^3) in an ascending order using an oblique rotation of 20° relative to the anterior commissure-posterior commissure line. The first five volumes were discarded to allow for signal equilibration.

SPM8 software (Wellcome Trust Centre for Neuroimaging, London) on Matlab 7.10 (MathWorks, Natick, MA) was used for fMRI data analysis. We conducted standard preprocessing, including realignment, normalization of each participant's functional data to the Montreal Neurological Institute EPI template and smoothing with a Gaussian kernel of 6 mm FWHM.

General linear models that accounted for low frequency drifts (cut-off period 128s) and temporal autocorrelations (AR(1)) were estimated for each participant. For the analysis of task properties (N = 94), the first-level model included three regressors for each session: LE videos, HE videos, and rating phase. The first-level model for assessing training-related changes ($n_{compassion}$ = 28,

n_{memory} = 30) was based on the same regressors for each session, extended to both time points (pre- and post-training). Realignment parameters were entered in the design matrix to further correct for residual effects of head motion. Regressors were convolved with a hemodynamic response function. The following linear contrasts were computed on the first-level for the analysis of task properties (N = 94): (1) effect of LE videos and (2) effect of HE videos. For the analysis of training-related changes (total n = 58), we computed the first-level contrasts: (1) difference post > pre training LE videos (∆LE videos) and (2) difference post > pre training HE videos (∆HE videos). First-level contrast images were smoothed with a Gaussian kernel of 8 mm FWHM and entered into second-level random effects group analyses. First, to assess task properties at pre-training, a repeated measures ANOVA with the within subject factor video type (LE and HE videos) was conducted. Second, training-related changes in BOLD responses were assessed in a 2 (∆LE and ∆HE; change per video type, within subject) x 2 (compassion and memory; training group, between subject) repeated measures ANOVA.

Parametric modulation of ratings was analyzed by setting up first-level models with the regressors video, trial-by-trial rating and rating phase separately for ratings of empathy, positive and negative affect. First level models were set up with the regressors LE videos, HE videos, the corresponding trial-by-trial ratings and rating phase. For the analysis of task properties (N = 94), we only modeled pre-training ratings, whereas training-related changes ($n_{compassion}$ = 28) were assessed by extending the model to both time points (pre- and post-training). Realignment parameters were entered in the design matrix to further correct for residual effects of head motion. Regressors were convolved with a hemodynamic response function. Linear contrasts were derived for each parametrically modulated rating at pre-training. For the analysis of training-induced changes, we computed the difference post > pre training across both video types for positive affect ratings, negative affect ratings and empathy. Contrast images from the first-level were smoothed with a Gaussian kernel of 8mm FWHM and entered into random effects group analyses. In second-level random group analyses, we first analyzed parametric modulation of BOLD responses at pre-training in separate one sample *t*-tests for each rating (N = 94). In addition, we computed the differences of parametrically modulated BOLD responses in a second-level repeated measures ANOVA with the within subject factor affect rating (3 levels: positive affect, negative affect and empathy, N = 94). The validation of the SoVT as a new measure of empathy for suffering was performed by entering the contrasts of parametric trial-by-trial ratings of HE videos in a second-level one-sample *t*-test. Training-related changes in

parametrically modulated BOLD contrasts in the compassion group (n = 28) were analyzed in one-sample *t*-tests.

Neuroimaging results are reported at $p < .05$, corrected for family wise errors (FWE) using Gaussian random fields theory and a cluster extent threshold of k = 5 voxels. When restricting the search volume to a priori defined ROIs, results are reported with a FWE corrected significance threshold of $P < 0.05$ using small volume correction (SVC). The validity of the SoVT as a measure of empathy was tested by constraining the analyses to results from a recent meta-analysis on the neural network underlying empathy for pain in cue- and picture-based paradigms (Lamm et al., 2011). In order to test for the robustness of compassion-induced training effects, we tested the effects of compassion training in independent ROIs based on two previously conducted studies (see above).

Validation of the SoVT. Based upon the results of a normative study with 265 participants, three parallel video sets (A, B and C) assembled out of an array of 342 videos were matched in terms of empathy, valence and arousal ratings (see above). The parallel nature of the three video sets and the difference between HE and LE videos were confirmed by pre-training ratings of a validation sample (Figure 3.1 A-B; total N = 94). A 2 x 3 repeated measures multivariate analysis of variance (MANOVA) with the factors video type (2 levels: HE and LE videos; within subject) and video set (3 levels: video sets A, B and C; between subject) and the dependent variables empathy, positive affect and negative affect ratings confirmed that in the fMRI experiment, there was no significant difference between video sets ($F(6,180) = 1.11$, $p = .36$), while the main effect of video type was significant ($F(3,89) = 397.78$, $p < .001$, $\eta^2 = 0.93$). The interaction video type x video set was not significant ($F(6,180) = 0.87$, $p = .52$). Follow-up *t*-tests confirmed that HE compared to LE videos evoked stronger empathy and negative affect ratings and less positive affect ratings (all $t(93) \geq 8.13$, all $p < .001$; Figure 3.2 A-C). In addition, we observed that all bivariate correlations between ratings of empathy, positive and negative affect were significant (all $r \geq .34$, all $p < .01$), indicating that before training, affective responses to the videos of the SoVT co-varied on all three dependent variables.

To determine the neural underpinnings of the SoVT before training, we conducted a second-level repeated measures ANOVA with the within subject factor video type (2 levels: activation when viewing HE videos and activation when viewing LE videos). HE compared to LE videos evoked stronger neural responses in an extended network comprising areas in occipital, temporal, frontal and parietal lobes, as well as subcortical areas including the amygdala. This

parallels previous findings using negative emotion pictures (Vrticka et al., 2011) or videos of negative valence (Goldin et al., 2008). Conversely, LE compared to HE videos elicited stronger activation in the posterior insula, the mOFC, the fusiform and the precentral gyrus ($p < .05$, FWE corrected; Supplementary Table S3.4). Parametric analyses with trial-by-trial subjective ratings ($p < .05$, FWE corrected; Supplementary Figure S3.2 C-E, Supplementary Table S3.5) revealed that negative and positive affect were associated with two distinct activation patterns: positive affect ratings correlated with activation in the mOFC, an area repeatedly linked to positive valuation and subjective experiences of pleasantness (Kringelbach and Berridge, 2009; Rangel and Hare, 2010; Grabenhorst and Rolls, 2011). Conversely, negative affect ratings correlated with activation in occipital, temporal, parietal and frontal cortices, as well as structures like the amygdala and the thalamus, again in line with previous results on the processing of negative pictures and videos (Goldin et al., 2008; Vrticka et al., 2011). In strong resemblance to negative affect ratings, empathy ratings correlated with a network spanning occipital, temporal and parietal cortices, as well as subcortical areas including the thalamus and the amygdala. Contrasting parametric regressors of ratings against each other in a second-level repeated measures ANOVA ($p < .05$, FWE corrected; Supplementary Table S3.5) revealed that empathy and negative affect ratings activated similar structures, while comparing positive affect to either empathy or negative affect revealed distinct mOFC activation. Conversely, comparing negative affect or empathy against positive affect revealed significantly higher brain-behavior covariation in parietal, occipital and temporal areas and subcortical regions including the amygdala.

Analysis of the Memory Task. To validate the effectiveness of the memory control training, we computed a 2 x 2 repeated measures MANOVA with the within subject factor time (pre- and post-training) and the between subject factor training group (compassion and memory; total n = 56). The dependent variables were the number of correctly remembered words and the number of words remembered in the correct position (Figure 3.3 D-E). We found significant main effects for time ($F(2,53) = 25.84$, $p < .001$, $\eta^2 = 0.49$) and group ($F(2,53) = 9.71$, $p < .001$, $\eta^2 = 0.27$), as well as a significant time x group interaction ($F(2,53) = 10.28$, $p < .001$, $\eta^2 = 0.28$). Univariate ANOVAs showed that all effects were significant for both dependent variables (all $F(1,54) \geq 4.25$, all $p < .05$). Follow-up independent *t*-tests revealed that, whereas the groups did not differ at pre-training, the memory group performed better than the compassion group at post-training in terms of correctly remembered words ($t(54) = 3.16$, $p < .01$) and words remembered in the correct position ($t(43.2) = 5.09$, $p < .001$). In addition, paired *t*-tests showed that the memory group significantly increased in the number of correctly remembered words ($t(28) = 4.88$, p

< .001) and words remembered in the correct position ($t(28)$ = 6.6, p < .001). The number of correctly remembered words also increased in the compassion group ($t(25)$ = 2.52, p < .05).

Supplementary Figures and Tables

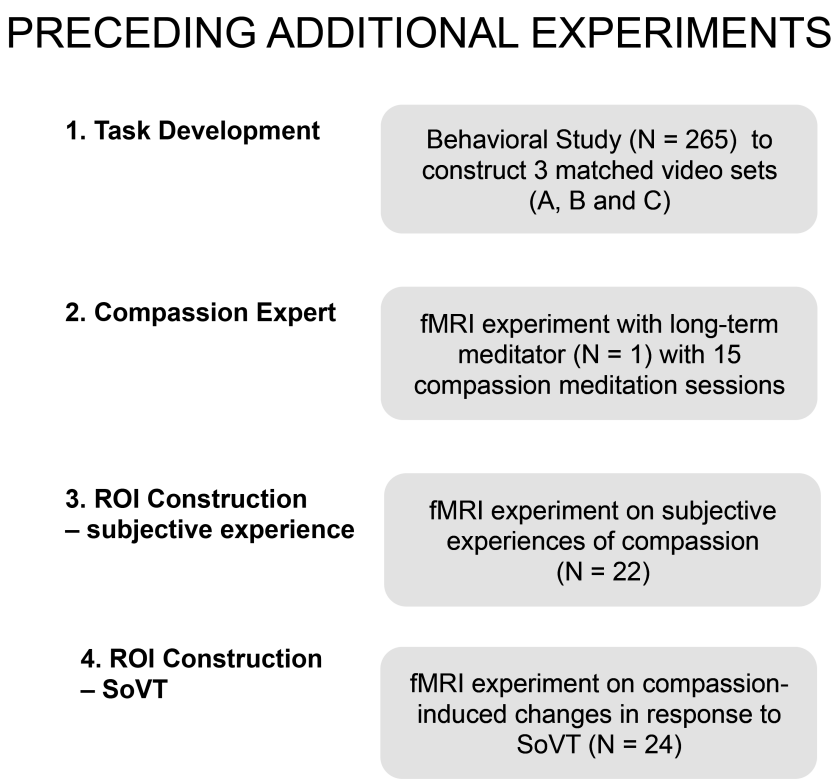


Figure S3.1. Preceding additional experiments. Behavioral study conducted to develop the SoVT (1). Neuroimaging studies conducted to assess neural networks underlying compassionate experiences (2-4). Detailed accounts of these experiments can be found in the Supplementary Text.

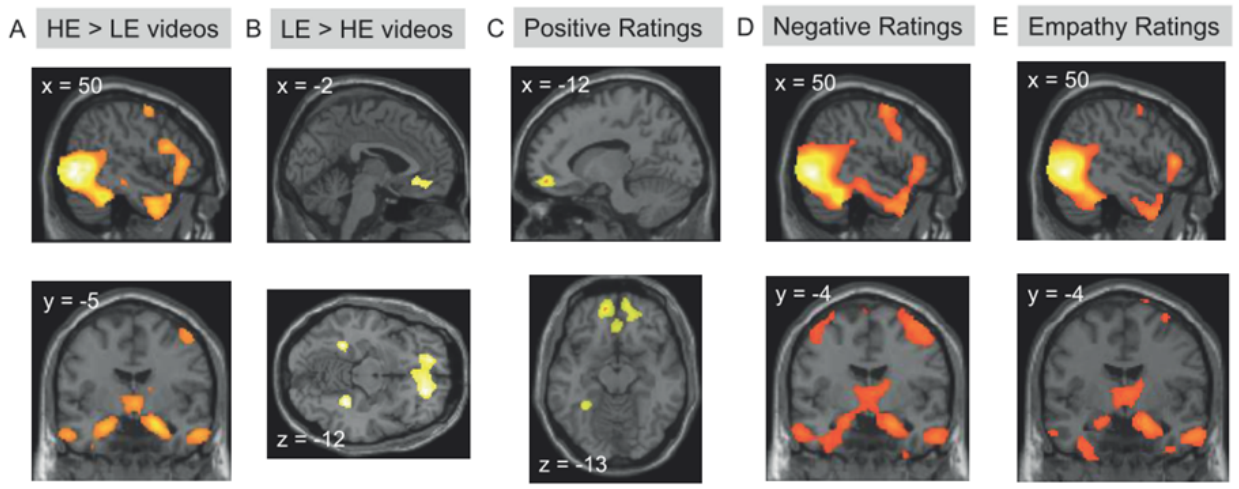


Figure S3.2. Neural correlates of the SoVT. (**A**) BOLD activations during viewing of high emotion compared to low emotion videos [HE > LE videos]. (**B**) BOLD activations during viewing of low emotion compared to high emotion videos [LE > HE videos]. (**C**) BOLD signal parametrically modulated by positive affect ratings (red activations at $p < .05$, FWE corrected; yellow activations at $p < .001$, uncorrected, shown for display purpose only). (**D**) BOLD signal parametrically modulated by negative affect ratings. (**E**) BOLD signal parametrically modulated by empathy ratings. All activations were rendered on an MNI template in neurological orientation. Apart from **C**, all activations are displayed at $p < .05$, FWE corrected, k = 5; color-coded, with brighter colors indicating lower p values. Inset x/y/z values indicate stereotactic coordinate of the shown slice in MNI space.

Table S1.1

Post-training debriefing questionnaire data from the compassion and the memory group

Question	Group	Mean	SD
How much did you like the training?	Memory	5.38	1.32
	Compassion	5.82	1.07
How well could you concentrate during the training?	Memory	5.72	0.87
	Compassion	5.00	0.91
How agreeable did you find the practices?	Memory	5.80	1.05
	Compassion	5.19	0.80
How easy did you find the practices?	Memory	6.08	0.87
	Compassion	4.87	0.92
How easy could you memorize your own route?/	Memory	6.02	1.24
How easy did you find Metta meditation for yourself?	Compassion	5.43	0.78
How easy could you memorize the route through Zurich?/	Memory	5.32	1.50
How easy did you find Metta meditation for a close person?	Compassion	5.76	0.88
How easy could you visualize the items to be remembered?/	Memory	5.88	0.84
How easy did you find Metta meditation for a neutral person?	Compassion	4.25	1.25
How easy could you link the items with your own locations?/	Memory	6.08	1.19
How easy did you find Metta meditation for a person in difficulty?	Compassion	5.27	1.22
How easy could you link the items with the locations in Zurich?/	Memory	5.54	1.13
How easy did you find Metta meditation for all persons?	Compassion	4.61	1.35

Note. Memory Group n = 30, Compassion Group n = 27. Ratings were obtained on a 7 point Likert scale from 1, not at all to 7, very well; SD, standard deviation.

Table S3.2

Qualitative self-reports of important or enjoyable experiences during the compassion training

The experience of how easy it is to feel Metta, given the right conditions and an open spirit.
How valuable silence is.
Silence does good. One can focus intensely on a single thought.
To perceive the changes in the surroundings before and after training. Afterwards, much more satisfied and, although tired, full of zest for action.
That it is possible in the first place to mentally focus so intensively.
Good, I enjoyed the effects enormously.
During the concentration I felt an enormous heat. It was nice to feel that one could create something like that inside oneself.
I was very much with myself. Never was so immersed in myself before. The silence was good for me.
A very pleasant feeling.
It was easier than I thought to generate feelings of benevolence towards a stranger.
It was very useful not to speak, it supported the effect.
To evoke positive feelings (almost) no matter when.
Was difficult to tune out sounds.
Since I am a person who is continuously in motion and never does anything to rest, it was a very positive experience for me. And to get away from everyday life.
I found a peaceful space in my head.
Practicing concentration for a longer period of time.
Relaxation, warmth.
The comfortable feeling, the unconditional nature of it.
Peace.
Walking meditation, silence, not speaking.
To get a break from daily routine.
To drift out.
Being together in silence.
Butterflies in the stomach, diving in, silence.
Good that there was variety (meditation inside and outside), there were enough breaks and the meditations did not last too long.
I was totally immersed.
To be turned inwards, the positive feelings, feelings of love, security, shelter.
Walking was very pleasant, I could concentrate better. It was exciting to feel the different intensities.
Good surrounding to contemplate things unrelated to daily routine.
The silence, being turned inward.

The warm feeling streaming through the body.
Peace, the feeling.
Silence, to explore oneself, contemplation of oneself.
The feeling of warmth, the feeling that it works.
Wonderful, fulfilling feeling to wish others well, pleasant teacher.
A feeling of happiness that arises that the teacher termed "quality of compassion".
Silence.
To think of other people and to really take the time to give them kindness.
That one could be for oneself without having to recount it to anyone.
Instructions, breathing exercises.

Table S3.3

Item Properties of the Socio-affective Video Task (SoVT).

Video Set	Video Type	N	Duration (s)	Valence Mean	Valence SD	Arousal Mean	Arousal SD	Empathy Mean	Empathy SD
A	HE	36	11.29	3.58	1.38	4.86	1.94	5.5	1.84
A	HE	36	10.41	3.78	1.15	4.14	1.97	5.31	2.08
A	HE	36	14.92	3.86	1.13	4.19	2.08	5.08	2.16
A	HE	33	11.37	3.42	1.06	4.18	2.16	6.12	1.45
A	LE	33	14.35	6.03	1.36	4.52	2.14	1.82	2.05
A	HE	33	10.31	3.64	1.11	5.12	2.21	5.64	1.45
A	LE	33	12.02	6.24	1.12	2.21	1.56	0.97	1.4
A	LE	34	11.74	5.35	1.28	2.85	2.25	0.5	1.21
A	HE	34	14.43	2.76	1.71	5.88	2.11	6.06	2.32
A	LE	35	10.17	6.57	1.44	2.03	1.6	0.89	1.26
A	LE	33	13.47	5.55	1.06	2.06	1.41	1.58	2.09
A	HE	33	11.74	2.88	1.65	5.7	2.4	5.79	2.16
A	HE	33	13.78	3.55	1.56	4.79	2.33	5.55	2.31
A	LE	33	13.23	4.97	1.4	2.91	2.24	2.36	2.33
A	HE	33	13.61	3.33	1.58	5.03	2.13	5.91	1.89
A	LE	33	12.58	5.12	1.58	3.15	2.31	2.36	2.67
A	LE	32	13.1	6	1.22	2.72	1.87	1.5	2.08
A	LE	32	12.3	5	1.14	3.06	2.12	2.56	2.51
A	HE	33	13.17	3.67	1.19	3.24	1.95	5.09	1.91
A	LE	33	11.18	5	1.25	2.88	2.16	0.76	1.28
A	LE	29	12.23	5.52	1.35	2.55	1.64	1.9	1.9
A	HE	29	10.91	3.62	1.52	3.76	2.47	4.9	2.32
A	LE	29	12.16	5.69	1.29	1.69	1.29	0.17	0.6
A	HE	29	10.02	4	1.34	3.69	2.42	4.76	2.31
B	HE	36	11.89	3.33	1.64	5.11	2.23	5.86	2.11
B	HE	36	14.34	4	1.31	4.39	2.23	4.69	2.15
B	HE	36	13.52	2.83	1.34	5.19	2.19	6.31	1.83
B	HE	33	10.81	3.15	1.23	5.97	1.99	5.36	1.45
B	HE	33	15.18	3.91	1.57	4.3	2.17	5.39	2.16

B	HE	33	14.85	4.15	1.35	3.58	2.05	4.67	2.23
B	LE	33	11.85	5	1.52	3.64	2.3	1.88	1.93
B	LE	34	14.91	5.29	0.91	1.94	1.49	0.68	1.32
B	LE	35	16.72	4.69	1.08	2.97	2.19	2.51	2.29
B	HE	33	13.05	3.88	1.45	3.76	2.02	5.27	1.83
B	LE	33	11.69	5.67	1.34	2.09	1.67	0.85	1.54
B	LE	33	12.81	6.36	1.6	2.09	1.91	0.48	1.35
B	LE	33	11.61	5.7	1.38	2.79	1.98	2.27	2.11
B	HE	33	10.31	4.03	1.69	4.24	2.46	5.21	2.12
B	LE	32	14.02	5.06	1.01	2.69	1.91	0.53	1.34
B	LE	32	12.58	5.28	0.89	2.19	1.4	1	1.74
B	LE	32	12.81	5.69	1.79	2.22	2.06	0.88	2.03
B	LE	32	17.73	5.41	0.98	3.28	2.19	1.75	1.93
B	HE	33	12.11	3.06	1.52	4.27	2.16	5.91	2.19
B	HE	33	14.04	3.55	1.5	3.45	2.31	5.58	2.22
B	LE	33	15.03	4.94	1.32	2.3	1.65	2.73	2.08
B	LE	33	14.01	4.7	1.16	2.76	1.66	2.61	2.29
B	HE	29	10.11	3.79	1.57	4.34	2.64	5.28	1.89
B	HE	29	10.59	3.45	1.18	4.59	2.54	5.59	1.99
C	HE	36	12.48	3.81	1.33	5.11	2.1	5.53	1.75
C	LE	36	14.96	4.89	0.4	2.69	2	1.11	1.67
C	LE	36	10.8	5.47	1.23	2.56	1.78	0.94	1.31
C	HE	36	12.02	3.31	1.31	4.61	2.14	5.94	2.04
C	LE	36	14.56	5.56	1	3.14	1.76	1.94	1.82
C	HE	36	11.28	3.31	1.39	4.97	2.24	5.56	2.1
C	HE	36	14.96	2.97	1.32	5.03	2.35	5.97	1.95
C	HE	36	14.64	3.56	1.4	4.89	2.23	5.42	2.27
C	LE	33	11.68	6.24	1.12	2.76	1.77	1	1.5
C	HE	33	12.02	3.15	1.12	5.33	2.01	5.85	1.84
C	LE	33	10.4	5.82	1.53	2.79	1.67	2.3	1.9
C	HE	33	11.28	3.12	1.36	5.33	2.42	5.7	1.98
C	LE	33	14.56	6.82	1.25	3.3	2.22	2.33	2.38
C	HE	34	11.12	3.94	1.43	3.79	2.31	4.97	2.1
C	HE	34	13.28	3.29	1.38	4.79	2.1	5.62	2.36
C	LE	34	10.32	5.71	2.01	2.47	1.73	1.71	2.32
C	LE	34	12.24	4.47	0.93	3.15	2.43	0.97	1.59
C	LE	34	14.56	6.62	1.13	2.59	1.86	0.53	1.21

C	LE	33	14.8	5.52	1.58	2.15	1.86	0.21	0.74
C	HE	33	14.96	3.12	1.29	5.03	2.51	5.52	1.87
C	HE	33	14.4	3.88	1.69	4.39	2.37	5.73	2.24
C	HE	33	10.45	3.15	1.7	4.94	2.4	6.15	1.91
C	LE	33	10.96	6.33	1.73	2.76	2.11	1.73	2.18
C	LE	32	13.44	4.91	0.82	1.88	1.43	0.44	1.22

Note. N, number of participants that rated each video item. LE, low emotion videos; HE, high emotion videos; SD, standard deviation.

Table S3.4

Foci of significant activation when viewing high emotion vs. low emotion videos [HE > LE videos], and vice versa [LE > HE videos] at pre-training measurement (N = 94)

Brain regions		Cluster size	MNI coordinates (mm)			Z values
			x	y	z	
HE > LE videos						
Inferior temporal gyrus	R	7070	48	-74	-8	>8.21
Middle temporal gyrus	R		52	-58	0	>8.21
Fusiform gyrus	R		46	-44	-28	>8.21
Cerebellum	L	13495	-16	-80	-46	>8.21
Inferior occipital gyrus	L		-48	-76	-8	>8.21
Lingual gyrus	L		-10	-68	-2	>8.21
Superior parietal gyrus	R	1229	34	-54	66	>8.21
Inferior frontal gyrus, pars triangularis	R	1355	60	32	-2	6.78
	R		50	18	20	5.84
Amygdala	R	3629	20	-4	-20	6.71
Superior temporal gyrus	L		-30	14	-32	6.42
Brainstem	M		0	-34	-6	6.23
Middle temporal gyrus	R	1659	60	2	-30	6.61
Inferior temporal gyrus	R		54	6	-34	6.53
Middle temporal gyrus	R		46	10	-44	6.34
Superior parietal gyrus	L	682	-34	-52	66	6.44
Middle frontal gyrus	R	139	50	0	58	5.49
Supramarginal gyrus	L	215	-58	-40	24	5.11
Superior frontal gyrus	R	185	10	14	72	5.1
	R	115	12	58	26	5.05
Caudate nucleus	R	61	16	2	12	4.9
Lingual gyrus	R	13	16	-62	-8	4.59
Inferior frontal gyrus, pars triangularis	L	12	-52	36	6	4.53

LE > HE videos						
Anterior orbital gyrus	R	991	20	44	-14	6.15
Medial orbital gyrus (mOFC)	L		-16	46	-12	5.7
Anterior cingulate cortex	M		0	32	-8	5.4
Fusiform gyrus	R	163	30	-42	-12	6.15
	L	97	-30	-44	-10	5.98
Insula	L	12	-40	-18	14	4.59
Precentral gyrus	R	12	4	-32	64	4.51

Note. All values $p < .05$, FWE corrected, spatial extent threshold of k = 5 voxels. M, middle; L, left; R, right; mOFC, medial orbitofrontal cortex; MNI coordinates and *Z* values are related to the peaks of the respective cluster. When clusters contain different subregions, these are specified without reiterating the overall cluster size.

Table S3.5

Foci of significant activation revealed by parametric modulation with empathy, positive affect and negative affect ratings (N = 94)

Brain regions		Cluster size	MNI coordinates (mm)			Z values
			x	y	z	
Positive Affect						
Medial orbital gyrus (mOFC)	L	5	-12	54	-12	4.53
Negative Affect						
Inferior temporal gyrus*	R	47675	52	-72	-6	>8.21
Inferior occipital gyrus	L		-50	-80	-8	>8.21
Inferior temporal gyrus	R		48	-46	-26	>8.21
Superior parietal gyrus	R	3314	36	-54	66	7.84
	L		-36	-52	70	6.69
Angular gyrus	R		34	-42	52	5.95
Superior frontal gyrus	R	795	10	58	22	5.5
	R		14	58	42	4.98
	L		-8	58	30	4.97
Fusiform gyrus	R	151	40	-12	-40	5.47
Supramarginal gyrus	L	715	-68	-26	28	5.28
	L		-66	-30	40	5.19
Precentral gyrus	L	157	-44	-4	58	4.8
Superior frontal gyrus	L		-34	-2	70	4.7
Precentral gyrus	L		-56	-2	48	4.53
Brainstem	R	22	8	-32	-30	4.61
Inferior frontal gyrus, pars triangularis	L	14	-42	14	24	4.54
Superior frontal gyrus	R	5	14	38	60	4.42
Empathy						
Inferior temporal gyrus	R	44360	52	-70	-4	>8.21
Inferior occipital gyrus	L		-52	-78	-8	>8.21

Lingual gyrus	L		-12	-72	-8	>8.21
Supramarginal gyrus	L	365	-70	-28	28	5.96
Supplementary motor area	R	106	16	2	80	4.9
	R		10	12	76	4.59
Middle frontal gyrus	R	53	52	0	54	4.8
Rectal gyrus	R	20	2	62	-20	4.55
Supplementary motor area	L	21	-8	0	80	4.55
Brainstem	R	12	2	-34	-34	4.55
Superior frontal gyrus	R	21	36	-4	68	4.54
Positive Affect > Empathy						
Anterior orbital gyrus	R	119	20	46	-16	4.75
Medial orbital gyrus (mOFC)	R		14	52	-10	4.59
	L	46	-14	50	-10	4.73
Anterior cingulate cortex	L	16	-2	32	-8	4.54
Positive Affect > Negative Affect						
Anterior orbital gyrus	R	95	22	48	-16	4.89
Anterior cingulate cortex	L	28	-2	32	-8	4.67
Medial orbital gyrus (mOFC)	L	10	-16	50	-8	4.4
Negative Affect > Positive Affect						
Inferior temporal gyrus	R	9882	52	-70	-6	>8.21
Cerebellum	R		48	-46	-30	7.58
Fusiform gyrus	R		44	-58	-22	7.5
Inferior occipital gyrus	L	10097	-48	-80	-10	>8.21
Cerebellum	L		-14	-82	-50	7.63
	L		-20	-78	-38	7.39
Superior parietal gyrus	R	791	38	-54	70	6.73
Amygdala	R	1865	18	-4	-20	6.64
Brainstem	R		10	-26	-10	5.7
Hypothalamus	L		-4	-2	-6	5.17
Inferior temporal gyrus	L	2366	-42	4	-36	6.53
	L		-32	16	-34	6.41
	L		-38	-6	-44	5.53
Middle occipital gyrus, superior part	R	355	28	-88	42	6.45

Middle frontal gyrus	R	678	52	0	56	6.18
	R		20	4	78	5.27
	R		12	14	74	4.9
Superior parietal gyrus	L	623	-36	-54	70	5.79
Amygdala	L	102	-20	-6	-20	5.38
Superior frontal gyrus	R	356	12	58	20	5.19
	R		16	50	50	4.88
	R		20	60	38	4.85
Superior frontal gyrus	L	53	-10	8	76	4.75
Inferior frontal gyrus, pars triangularis	L	28	-44	14	22	4.73
Supramarginal gyrus	L	332	-60	-30	38	4.73
	L		-68	-36	32	4.61
	L		-68	-24	28	4.56
Precentral gyrus	L	32	-42	-4	56	4.43
Thalamus	R	12	6	-16	14	4.41
	R	7	18	-16	14	4.4
Inferior temporal gyrus	R	8	34	-6	-48	4.38
Negative Affect > Empathy						
n.s.						
Empathy > Positive Affect						
Inferior temporal gyrus	R	6261	52	-70	-6	>8.21
Fusiform gyrus	R		44	-58	-20	>8.21
Cerebellum	R		48	-46	-30	7.3
Inferior occipital gyrus	L	13562	-48	-78	-8	>8.21
Cerebellum	L		-12	-82	-48	7.74
	L		-44	-44	-30	7.46
Amygdala	R	5628	20	-2	-22	7.02
Inferior frontal gyrus, pars triangularis	R		48	34	6	6.75
Middle temporal gyrus	R		58	2	-32	6.56
Superior parietal gyrus	R	1113	40	-50	70	7.01
Superior occipital cortex	R	457	28	-88	44	6.87
Superior parietal gyrus	L	838	-36	-54	70	6.7
Supramarginal gyrus	R	472	66	-28	28	6.05
Middle frontal gyrus	R	473	52	0	56	5.92

Middle frontal gyrus, superior part	R		20	4	78	5.54
	R		32	0	72	4.48
Supramarginal gyrus	L	406	-60	-30	38	5.07
	L		-68	-24	28	5
Superior frontal gyrus	R	49	12	58	20	4.98
Middle frontal gyrus, superior part	L	19	-12	8	78	4.44
Inferior frontal gyrus, p. opercularis	L	6	-44	14	22	4.42
Empathy > Negative Affect						
n.s.						
SVC: Empathy HE videos						
Brainstem	R		2	-30	-2	5.61
Thalamus	L		-8	-24	10	4.52
	L		-6	-20	6	4.42
Middle occipital gyrus	R		36	-88	8	4.82
Cingulate gyrus (aMCC)	M		0	10	34	4.66
	R		2	18	30	4.53
Insula (AI)	R		34	26	2	4.14
Putamen	R		28	20	2	3.73
Inferior frontal gyrus, pars orbitalis	R		44	26	-4	3.48
	L		-30	18	-22	3.94

Note. All values $p < .05$, FWE corrected, spatial extent threshold k = 5 voxels. Small volume correction (SVC) in regions of interest (Lamm et al., 2011) at $p < .05$, corrected for FWE. M, middle; L, left; R, right; mOFC, medial orbitofrontal cortex; aMCC, anterior medial cingulate cortex; AI, anterior insula. MNI coordinates and *Z* values are related to the peaks of the respective cluster. When clusters contain different subregions, these are specified without reiterating the overall cluster size. * this cluster encompasses the amygdala and the thalamus.

Table S3.6

Foci of activation in one expert meditation practitioner when immersing himself in three different compassionate states (loving kindness, unconditional compassion, pain compassion) to a high compared with a low degree

Brain regions		Cluster size	MNI coordinates (mm)			Z values
			x	y	z	
High > Low						
Caudate nucleus*	L	7825	-9	-4	13	>8.21
	R		12	-1	13	>8.21
	R		15	-7	19	>8.21
Supplementary motor area	L	1059	-12	5	70	>8.21
Middle cingulate gyrus	R		9	17	37	>8.21
Supplementary motor area	R		6	5	70	>8.21
Precentral gyrus	R	312	51	2	46	>8.21
	R		57	-4	37	>8.21
	R		54	5	34	>8.21
Inferior frontal gyrus, pars orbitalis	R	12	51	47	-11	>8.21
Middle occipital gyrus	L	93	-33	-70	22	>8.21
	L		-48	-82	19	7.23
	L		-45	-82	31	5.97
Inferior parietal gyrus	L	265	-57	-46	40	>8.21
	L		-66	-37	22	>8.21
Supramarginal gyrus	L		-66	-25	22	7.11
Superior parietal gyrus	M	86	0	-43	64	>8.21
	M		0	-52	70	>8.21
Superior parietal gyrus	L		-6	-64	67	5.72
Middle occipital gyrus	R	18	48	-82	19	>8.21
	R		54	-73	22	7.19
Middle temporal gyrus	R	22	66	-52	10	>8.21
	L	43	-42	8	-38	>8.21
Inferior temporal gyrus	L		-39	20	-35	7
Superior occipital gyrus	L	50	-12	-100	28	>8.21
	L		-15	-106	16	7.48

Middle occipital gyrus	L		-27	-100	13	7.17
Inferior temporal gyrus	R	60	48	-64	-8	7.84
	R		60	-64	-5	6.2
Middle occipital gyrus	R	61	30	-97	16	7.79
	R		18	-106	10	7.04
	R		33	-91	22	6.99
Enthorinal cortex	R	12	24	-7	-41	7.77
Inferior temporal gyrus	L	12	-57	-22	-35	7.7
Postcentral gyrus	R	57	33	-31	58	7.55
Middle orbital gyrus	R	14	9	44	-5	7.16
Middle temporal gyrus	L	39	-57	-70	10	7.05
	L		-60	-55	-2	6.55
	L		-60	-64	1	6.24
	R	13	54	-43	-2	7.02
Middle cingulate gyrus	M	24	0	-19	43	6.92
	M		0	-10	37	5.37
Superior frontal gyrus	R	12	15	35	55	6.86
Postcentral gyrus	L	21	-21	-31	61	6.57
Inferior parietal gyrus	L	13	-27	-46	49	6.05
Brainstem	R	11	6	-40	-44	6.04
Cerebellum	R		12	-49	-41	5.89
Precuneus	M	10	0	-58	19	5.89
Supramarginal gyrus	R	10	54	-34	31	5.62
Inferior parietal gyrus	R	10	36	-40	52	5.43

Note. All values $p < .05$, FWE corrected, spatial extent threshold of k = 10 voxels. M, middle; L, left; R, right; mOFC, medial orbitofrontal cortex. MNI coordinates and *Z* values are related to the peaks of the respective cluster. When clusters contain different subregions, these are specified without reiterating the overall cluster size. * this cluster encompasses the putamen, the pallidum, the ventral tegmental area/ substantia nigra and the mOFC as displayed in Figure 3.4.

Table S3.7

Regions of interest extracted from two independent compassion training studies (N = 46)

Region of Interest (MNI)		Volume (mm^3)	Centre of mass (MNI space)		
			x	y	z
Amygdala	L	872	-20,6	-2,3	-16
	R	864	24,6	0	-15,8
Middle frontal gyrus	R	1128	44,6	-2,5	56,6
Superior frontal gyrus	R	1376	22,6	-3,6	72,5
Fusiform gyrus	L	5880	-35,6	-48,3	-19,6
	R	1144	39,3	-21,3	-25,9
Hippocampus	R	904	37	-17,1	-15,2
Lingual Gyrus	L	1000	-19	-53,7	-5,8
Inferior occipital gyrus	L	2736	-46,8	-71,2	-11,2
Middle occipital gyrus	L	3360	-42,3	-74,6	12,1
Pallidum	L	1072	-19,6	-1,1	0,5
	R	1376	22,4	-3	-1,9
Parahippocampus	R	2952	23,1	-4,1	-26,8
Precentral gyrus	L	1840	-46,9	-1,8	51,3
	R	1632	36,2	-8,8	59,9
Supramarginal Gyrus	L	1728	-57,7	-39,3	24,9
Putamen	L	4728	-22,9	5,9	1,1
	R	3168	28	-1,9	-2,5
Gyrus rectus (mOFC)	L	1544	-5,7	47,2	-18,4
	R	1752	8,5	40,5	-17,8
Inferior temporal gyrus	L	2312	-46,6	-50,9	-16,9
Middle temporal gyrus	L	6128	-50,8	-53,6	9
Superior temporal gyrus	L	5240	-56,3	-39,2	15,7
Thalamus	L	2120	-10,5	-13,7	9,1
	R	3776	12,1	-13,8	5,5
Ventral tegmental area/ substantia nigra (VTA/SN)	R	936	8,8	-16,3	-12,1

Note. ROIs were required to lie within the MNI regions or the VTA/SN (Filbey et al., 2007); $p < .005$, uncorrected; cluster extent > 810 mm^3; mOFC, medial orbitofrontal cortex.

Table S3.8

Foci of significant activation associated with differences in training-induced changes (Δ) between the compassion and memory group

Brain regions		Cluster size	MNI coordinates (mm) x	y	z	Z values
(Compassion ΔHE & ΔLE) > (Memory ΔHE & ΔLE)						
Inferior occipital gyrus	R	25	26	-106	-12	3.43
	L	22	-28	-108	-10	3.33
SVC: (Compassion ΔHE & ΔLE) > (Memory ΔHE & ΔLE)						
Putamen	R		28	-6	6	3.08
Pallidum	R		26	-8	4	2.88
(Memory ΔHE & ΔLE) > (Compassion ΔHE & ΔLE)						
Fusiform gyrus	L	327	-20	-82	-10	4.24
Hypothalamus	R	25	4	0	-6	3.53
Corpus callosum	R	51	18	-40	26	3.5
Lingual gyrus	R	41	14	-80	-6	3.36
SVC: (Memory ΔHE & ΔLE) > (Compassion ΔHE & ΔLE)						
n.s.						
(Compassion ΔHE- ΔLE) > (Memory ΔHE- ΔLE)						
Lateral orbital gyrus	R	885	46	56	-12	4.54*
	R		38	48	-18	4.33
Superior frontal gyrus	R	383	24	26	64	4.31
Angular gyrus	R	945	50	-60	44	4.05
	R		36	-74	46	3.86
Angular gyrus	L	146	-42	-78	46	4
Middle temporal gyrus	L	296	-66	-42	-14	3.71
	L		-68	-30	-14	3.57

Cerebellum	L	209	-40	-70	-42	3.63
Middle temporal gyrus	R	117	64	-42	-18	3.56
Superior frontal gyrus	L	40	-16	30	34	3.52
Inferior occipital gyrus	R	37	24	-100	-12	3.39
Inferior frontal gyrus, pars triangularis	L	20	-52	42	10	3.39
Superior parietal gyrus	R	29	20	-66	64	3.3
Caudate nucleus	L	7	-18	6	24	3.25
Cerebellum	R	17	34	-72	-32	3.24
Postcentral gyrus	L	13	-48	-22	64	3.23
Middle frontal gyrus	R	11	26	36	34	3.19
Fusiform gyrus	L	5	-28	-16	-38	3.16
SVC: (Compassion ΔHE - ΔLE) > (Memory ΔHE - ΔLE)						
Gyrus rectus (mOFC)	R		12	38	-18	2.73
(Memory ΔHE- ΔLE) > (Compassion ΔHE- ΔLE)						
n.s.						
SVC: (Memory ΔHE - ΔLE) > (Compassion ΔHE - ΔLE)						
n.s.						
Compassion ΔHE > Memory ΔHE						
Inferior occipital gyrus	R	160	26	-106	-12	4.16
	L	87	-28	-106	-8	3.79
Inferior frontal gyrus, pars triangularis	R	68	34	26	28	3.56
Medial orbital gyrus	R	30	18	38	-16	3.3
Putamen	R	19	26	-6	4	3.27
SVC: Compassion ΔHE > Memory ΔHE						
Putamen	R		26	-6	4	3.27
Pallidum	R		26	-8	4	3.24
Gyrus rectus (mOFC)	R		14	38	-16	2.89
Ventral tegmental area/ substantia nigra (VTA/SN)	R		4	-16	-12	2.46
	R		12	-14	-12	2.45
Memory ΔHE > Compassion ΔHE						
Fusiform Gyrus	L	217	-20	-82	-10	4

SVC: Memory ΔHE > Compassion ΔHE

n.s.

Compassion ΔLE > Memory ΔLE

n.s.

SVC: Compassion ΔLE > Memory ΔLE

n.s.

Memory ΔLE > Compassion ΔLE

Lingual gyrus	R	317	12	-78	-6	3.95
Fusiform gyrus	R		26	-78	-12	3.36
	L	199	-20	-82	-10	3.87
Middle temporal gyrus	L	111	-58	-44	-12	3.82
Hippocampus	L	92	-32	-22	-16	3.64
Hypothalamus	R	30	6	0	-6	3.57
Corpus callosum	R	63	18	-40	26	3.46
Precentral gyrus	R	50	14	-32	70	3.46
	M	13	0	-28	64	3.22

SVC: Memory ΔLE > Compassion ΔLE

Hippocampus	R		40	-16	-22	2.57
Fusiform gyrus	R		40	-16	-24	2.71

Note. Values $p < .001$, uncorrected, spatial extent threshold k = 5 voxels. Small volume correction (SVC) in ROIs (Table S3.5) at $p < .05$, FWE corrected. * denotes $p < .05$, FWE corrected. M, middle; L, left; R, right; mOFC, medial orbitofrontal cortex. MNI coordinates and *Z* values are related to the peaks of the respective cluster. When clusters contain different subregions, these are specified without reiterating the overall cluster size.

Table S3.9

Sample characteristics for participants in compassion and memory training

Variable	**Compassion (n = 28)** Mean (SD)	**Memory (n = 30)** Mean (SD)
Age (years)	24.39 (3.56)	22.7 (3.95)
Highest Completed Education	1 secondary school or apprenticeship 16 high school 11 university or higher	3 secondary school or apprenticeship 22 high school 5 university or higher
Alexithymia[1]	41.54 (7.27)	40.37 (7.21)
Depression[2]	5.21 (3.35)	3.37 (3.66)
Fantasy[3]	17.54 (4.73)	15.67 (4.84)
Empathic Concern[3]	20.18 (3.55)	20.5 (3.15)
Personal Distress[3]	11.64 (5.01)	10.33 (3.17)
Perspective Taking[3]	18.75 (4.99)	19.90 (3.23)
Compassionate Love[4]	88.07 (22.8)	89.4 (16.76)

Note. The compassion and memory sample did not differ on the socio-demographic and questionnaire variables: [1]Toronto Alexithymia Scale (TAS (Bagby et al., 1994); > 60 clinically relevant); [2]Beck's Depression Inventory (BDI (Beck et al., 1996); > 18 clinically relevant); [3]Subscales of Interpersonal Reactivity Index (IRI (Davis, 1983)); [4]Compassionate Love Scale (CLS (Sprecher and Fehr, 2005)).

Table S3.10

Sample characteristics for participants in validation sample (total N = 94) viewing different video sets

Variable	Video Set A (n = 33)	Video Set B (n = 31)	Video Set C (n = 30)
	Mean (SD)	Mean (SD)	Mean (SD)
Age (years)	24.7 (3.99)	24.94 (4.46)	23.2 (3.98)
Highest Completed Education	1 secondary school or apprenticeship 16 high school 16 university or higher	3 secondary school or apprenticeship 20 high school 8 university or higher	3 secondary school or apprenticeship 21 high school 6 university or higher
Alexithymia[1]	40.46 (7.75)	42.26 (8.15)	40.3 (7.38)
Depression[2]	4.7 (4.61)	5.94 (4.21)	4.37 (3.95)
Fantasy[3]	16.39 (4.96)	17.03 (4.82)	16.37 (4.69)
Empathic Concern[3]	19.42 (3.66)	21.19 (3.91)	20.73 (2.95)
Personal Distress[3]	11.61 (3.67)	12.07 (4.28)	11.1 (4.13)
Perspective Taking[3]	19.39 (4.09)	19.65 (3.84)	18.73 (4.47)
Compassionate Love[4]	92.61 (19.32)	90.36 (19.35)	83.33 (21.68)

Note. The groups viewing different video sets did not differ on the socio-demographic and questionnaire variables: [1]Toronto Alexithymia Scale (TAS (Bagby et al., 1994); > 60 clinically relevant); [2]Beck's Depression Inventory (BDI (Beck et al., 1996); > 18 clinically relevant); [3]Subscales of Interpersonal Reactivity Index (IRI (Davis, 1983)); [4]Compassionate Love Scale (CLS (Sprecher and Fehr, 2005)).

4. Differential Pattern of Functional Brain Plasticity after Compassion and Empathy Training

Corresponding publication:

Klimecki, O., Leiberg, S., Ricard, M., Singer, T. (2013). Differential Pattern of Functional Brain Plasticity after Compassion and Empathy Training. Soc Cogn Affect Neurosci. doi: 10.1093/scan/nst060. By permission of Oxford University Press.

Abstract

Although empathy is crucial for successful social interactions, excessive sharing of others' negative emotions may be maladaptive and constitute a source of burnout. To investigate functional neural plasticity underlying the augmentation of empathy and to test the counteracting potential of compassion, one group of participants was first trained in empathic resonance and subsequently in compassion. In response to videos depicting human suffering, empathy training, but not memory training (control group), increased negative affect and brain activations in anterior insula and anterior midcingulate cortex – brain regions previously associated with empathy for pain. In contrast, subsequent compassion training could reverse the increase in negative effect and, by contrast, augment self-reports of positive affect. In addition, compassion training increased activations in a non-overlapping brain network spanning ventral striatum, pregenual anterior cingulate cortex, and medial orbitofrontal cortex. We conclude that training compassion may reflect a new coping strategy to overcome empathic distress and strengthen resilience.

Our capacity to understand others' feelings through empathy is crucial for successful social interactions (Bird et al., 2010). However, when confronting the suffering of others, intense sharing of the other's pain can be a primary cause for empathic distress and decreased helping behavior (Batson et al., 1987; Eisenberg et al., 1989). In fact, empathic responses to witnessing another in pain are usually experienced as aversive (Lamm et al., 2011). This may be especially problematic for people working in professions where suffering is routinely encountered. Physicians, for example, have a high prevalence rate of burnout (McCray et al., 2008) and an elevated risk for suicide (Schernhammer and Colditz, 2004). A potential remedy for the excessive sharing of negative affect may be compassion. Compassion is defined as a feeling of concern for the suffering of others that is associated with the motivation to help (Keltner and Goetz, 2007). Recent studies of others and ourselves have shown that training compassion can foster emotional well-being (Fredrickson et al., 2008), positive emotions (Klimecki et al., 2012) and prosocial behavior (Leiberg et al., 2011). Although compassion emerges as a promising strategy to strengthen personal resources, it is, so far, unresolved how compassion can help to overcome the adverse effects related to empathic distress.

Furthermore, it is unresolved whether the neural systems subserving empathy and compassion can be dissociated. Two recent cross-sectional meta-analyses suggest that empathy for pain crucially involves anterior insula (AI) and anterior midcingulate cortex (aMCC) (Fan et al., 2011; Lamm et al., 2011). This is consistent with the observation that negative affect often covaries with activations in AI and aMCC (Lamm et al., 2011). On a more general level, AI and aMCC are key structures for processing salient events (Seeley et al., 2007) and aMCC function has been robustly implicated in cognitive control and pain processing in two recent large-scale meta-analyses (Beckmann et al., 2009; Shackman et al., 2011). Conversely, several cross-sectional studies (Beauregard et al., 2009; Immordino-Yang et al., 2009; Kim et al., 2009; Lutz et al., 2008) and one short-time longitudinal study performed by our group (Klimecki et al., 2012)

suggest that compassion is accompanied by activations in regions typically associated with reward, love, and affiliation. These regions comprise insula, ventral striatum, and medial orbitofrontal cortex (mOFC) (Beauregard et al., 2009; Immordino-Yang et al., 2009). Animal studies suggest that the neurobiology of the "care" system can be clearly dissociated from other emotional-motivational systems such as the "panic" system, as it relies on distinct brain structures and is mediated by distinct neurotransmitters comprising opioids, oxytocin and dopamine (Panksepp, 2011). In addition, affiliative memories in mammals seem to rely on a circuitry that includes mOFC, ventral striatum, and ventral tegmental area (Depue and Morrone-Strupinsky, 2005). Our aim was thus to determine whether training empathy and compassion will have distinct effects on neural function and whether training compassion can help overcome excessive levels of distress.

To address these issues, we conducted a prospective training study in which one group of participants was first trained in empathy and subsequently in compassion. We repeatedly acquired functional magnetic imaging (fMRI) measures while participants were exposed to videos depicting others suffering. To train compassion, we used a contemplative technique from secular compassion training programs that aims at cultivating feelings of benevolence and friendliness in a state of quiet concentration (Salzberg, 2002; for empirical work, see Fredrickson et al., 2008; Leiberg et al., 2011; Klimecki et al., 2012). Similar to strengthening modes of affiliation, compassion training relies on extending caring feelings usually experienced towards close loved persons to other human beings. The preceding empathy training closely matched the compassion training in form and structure, but focused on resonating with suffering. Unspecific effects introduced by training in groups and by repeated measurements were controlled by including an active control group that received memory training using the Method of Loci (Bower, 1970). A detailed description of the employed training techniques can be found in the Supplementary Material.

On the level of subjective experience, we hypothesized that training empathy would increase empathy and negative affect when witnessing the distress of others. Pertaining to neural function, we assumed that training empathy would induce plasticity in AI and aMCC as these structures are robustly involved in cross-sectional studies on empathy for pain (Fan et al., 2011; Lamm et al., 2011). In contrast, we expected that a subsequent compassion training would strengthen positive affect and induce specific functional plasticity in a different neural network. This network includes mOFC, ventral tegmental area/substantia nigra (VTA/SN), and striatum, as compassion-related activation changes in these structures have been observed in our recent longitudinal study (Klimecki et al., 2012).

METHODS

Participants. As gender differences in social emotions were observed in previous neuroscientific research (e.g. Singer et al., 2006), we decided to control for possible gender effects by restricting our sample to female participants only. In the affect group, the study was completed by 25 of an initial 30 participants (age: 25.88 ± 4.32 years, mean ± SD). In the memory group 28 of 33 participants completed the study (age: 22.89 ± 4.02 years, mean ± SD). Participants for the affect and memory training groups (Figure 4.1) were recruited and tested sequentially due to temporal and infrastructural constraints (i.e., scanning slots). To avoid any selection bias, participants in both training groups were recruited with advertisements announcing participation in mental training studies. Furthermore, participants were not aware of the specific training content until pretest measurement was completed and they entered the training phase. The five persons who dropped out in the memory control group had higher scores on the Beck's Depression Inventory (Beck et al., 1996) than those who completed the study ($t(31) = 2.31$, $p < .05$, dropouts: mean = 7.4, SD = 5.6; completers: mean = 3.11, SD = 3.49). No other selective dropouts were observed in the memory control group and no selective

dropouts occurred in the affect group. To account for selective dropout in the memory group and age differences between both groups (Table S4.1), we included age and depression scores as covariates in all between group analyses. The study was approved by the Research Ethics Committee of Zurich (‘‘Kantonale Ethikkommission des Kantons Zürich – Spezialisierte Unterkommission Psychiatrie, Neurologie, Neurochirurgie’’; E-25/2008) and was carried out in compliance with the Declaration of Helsinki. All subjects gave informed written consent, were paid for their participation, and were debriefed after the completion of the study. As participants whose data are reported here were part of a larger study, we specify the relation between the present study and other experiments (Klimecki et al., 2012) in Figure S4.1. A description of inclusion criteria, the employed trait questionnaires, data acquisition and data analysis procedures, as well as the training regimes can be found in the Supplementary Material.

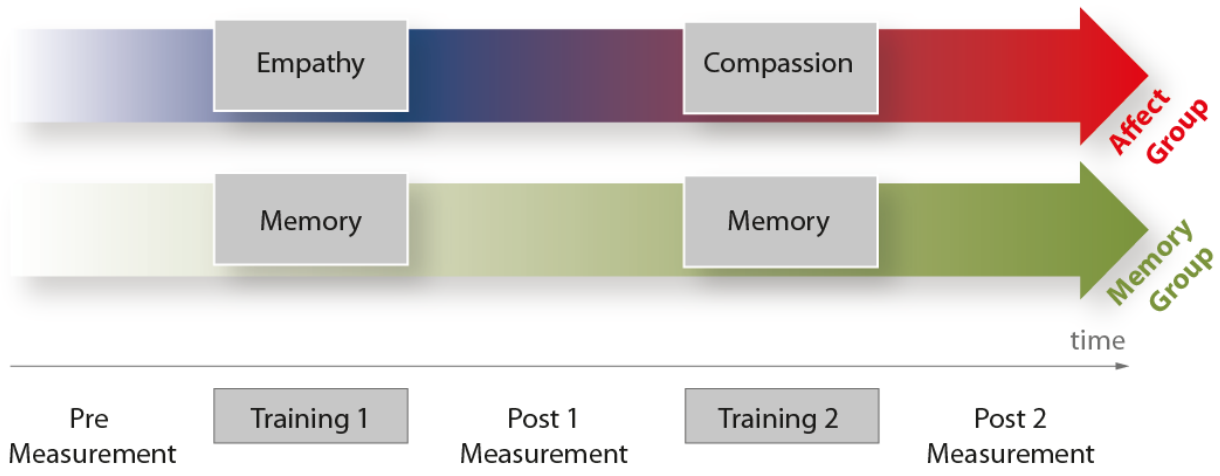


Figure 4.1. Experimental design. This longitudinal training study consisted of two groups: the affect group, which first received empathy training and subsequently compassion training, and the memory control group, which received two memory trainings. Participants were tested three times while watching videos depicting others suffering: Before the first training (Pre) and after each training (Post1 and Post2). Details on the training regimes can be found in the Supplementary Material.

Measures. Socio-affective Video Task (SoVT). Participants' affective experiences and blood oxygenation level dependent (BOLD) signals were measured three times in response to the SoVT (for more details about the SoVT properties please see also Klimecki et al., 2012) – before training (Pre), after empathy or memory training (Post1), and after compassion or memory training (Post2). To avoid habituation and repetition, participants saw one of three parallel video sets matched for valence, arousal and empathy at Pre, Post1, and Post2. Each set contained 12 high emotion (HE) videos and 12 low emotion (LE) videos. Video scenes were taken from footage cast for news or documentaries and depict men, women, and children. LE videos showed everyday scenes, while HE videos depicted people who were suffering (e.g., due to injuries or natural disasters). After each video (duration 10-18s) participants rated how much empathy, positive affect, and negative affect they had experienced while seeing the video. In order to assure that all participants had the same basic notion of empathy, they were instructed before each measurement that the empathy rating captures how much they shared the emotion of the depicted persons. Videos were shown in blocks of three HE or LE videos. Each block was followed by a null-event (10s fixation cross). At Post1 and Post2, participants in the affect training group were encouraged to make use of the trained competences when viewing the videos.

Memory Task. To test the effectiveness of the memory intervention, participants were seated in front of a computer screen at each measurement point and asked to encode a different matched list of 34 words. Words were presented for 4s each, followed by a 2s fixation cross. Subsequently, participants were given 5 min to fill the recollected words into a computer table, if possible in the correct sequence.

RESULTS

Socio-affective Video Task. To determine how the different training regimes affected subjective experiences of empathy, positive affect, and negative affect in response to the SoVT, we conducted a repeated measures multivariate analysis of variance (MANOVA). Age and depression scores (Beck et al., 1996) were included as covariates to account for selective dropout and between-group differences. The within subject factors were time (3 levels: Pre, Post1 and Post2) and video type (2 levels: LE and HE videos). The between subject factor was group (2 levels: affect and memory). The main effect of video type was significant ($F(3,47)$ = 8.18, $p < .001$, $\eta^2 = .34$). Significant interactions were observed for video type x group ($F(3,47)$ = 2.85, $p < .05$, $\eta^2 = .15$) and time x group ($F(6,44)$ = 6.01, $p < .001$, $\eta^2 = .45$). The triple interaction time x video type x group was marginally significant ($F(6,44)$ = 2.3, $p = .05$, $\eta^2 = .24$). Univariate ANOVAs determined that all three affect ratings showed the main effect of video type (all $F \geq 5.28$, all $p < .05$). Follow-up paired *t*-tests showed that compared to LE videos, HE videos elicited more negative affect and empathy and less positive affect (all $t(52) \geq 15.82$, all $p < .001$). The interaction video type x group was significant for empathy ($F(1,49)$ = 7.01, $p < .05$). Follow-up independent *t*-tests revealed that when combining all three time points, empathy ratings for LE videos were higher in the affect group than in the memory group ($t(51)$ = 3.68, $p < .01$). Importantly, the time x group interaction was significant for all three affect ratings (all $F \geq 3.44$, all $p < .05$). Follow-up independent *t*-tests comparing the memory and the affect groups at Pre- and Posttests showed that the groups did not differ before training (all $t \leq .29$, all $p \geq .77$), that empathy was higher in the affect group compared to the memory group after both trainings (both $t(51) \geq 3.18$, both $p < .01$), that a similar trend was present for negative affect after empathy training ($t(51)$ = 1.75, $p = .09$), and that positive affect was higher after compassion training ($t(51)$ = 4.37, $p < .001$). Paired *t*-tests focusing on changes within the affect group between Pre and Post1 and between Post1 and Post2, showed that empathy training increased negative affect ($t(24)$ = 3.5, $p < .01$) and empathy ($t(24)$ = 4.66, $p < .001$). Conversely,

compassion training decreased negative affect ($t(24)$ = 3.04, p < .01) and augmented positive affect ($t(24)$ = 4.25, p < .001). Paired t-tests in the memory group showed that negative affect decreased from Post1 to Post2 ($t(27)$ = 3.17, p < .01). Finally, Pearson correlations between the change in affect ratings and self-reports of practice duration in the affect group revealed no significant relation (all $p \geq .1$). In summary, training empathy led to increases in subjective reports of negative affect and empathy. Adding compassion training strengthened positive affect and reversed the observed increase in negative affect (Figure 4.2A).

Memory Task. To validate the effectiveness of the memory control training, we computed a 3 x 2 repeated measures MANOVA with the within subject factor time (3 levels: Pre, Post1 and Post2) and the between subject factor training group (2 levels: affect training and memory training group). Age and depression scores (Beck et al., 1996) were included as covariates. The dependent variables were the number of correctly remembered words and the number of words remembered in the correct position. We found a significant main effect of group ($F(2,46)$ = 11.35, p < .001, η^2 = .33) and a significant time x group interaction ($F(4,44)$ = 11, p < .001, η^2 = .5). There was a trend for a main effect of time ($F(4,44)$ = 2.23, p = .08). The time x group interaction was significant for both dependent variables (both $F(2,94) \geq 4.67$, both p < .05). Confirming the effectiveness of the memory intervention (n = 51; Figure 4.2B), follow-up independent t-tests revealed that, whereas the groups did not differ at Pre (both $p \geq .4$), the memory group performed better than the affect group at Post1 and at Post2 on both dependent measures (all $t \geq 3.43$, all p < .01). Paired t-tests showed that memory performance did not change in the affect group (all $t \leq 1.35$, all $p \geq .19$). In contrast, the number of correctly remembered words and words remembered in the correct position increased significantly in participants of the memory group, both, from Pre to Post1 (both $t \geq 4.37$, both p < .001) and from Post1 to Post2 (both $t \geq 2.35$, both p < .05). All other effects were not significant.

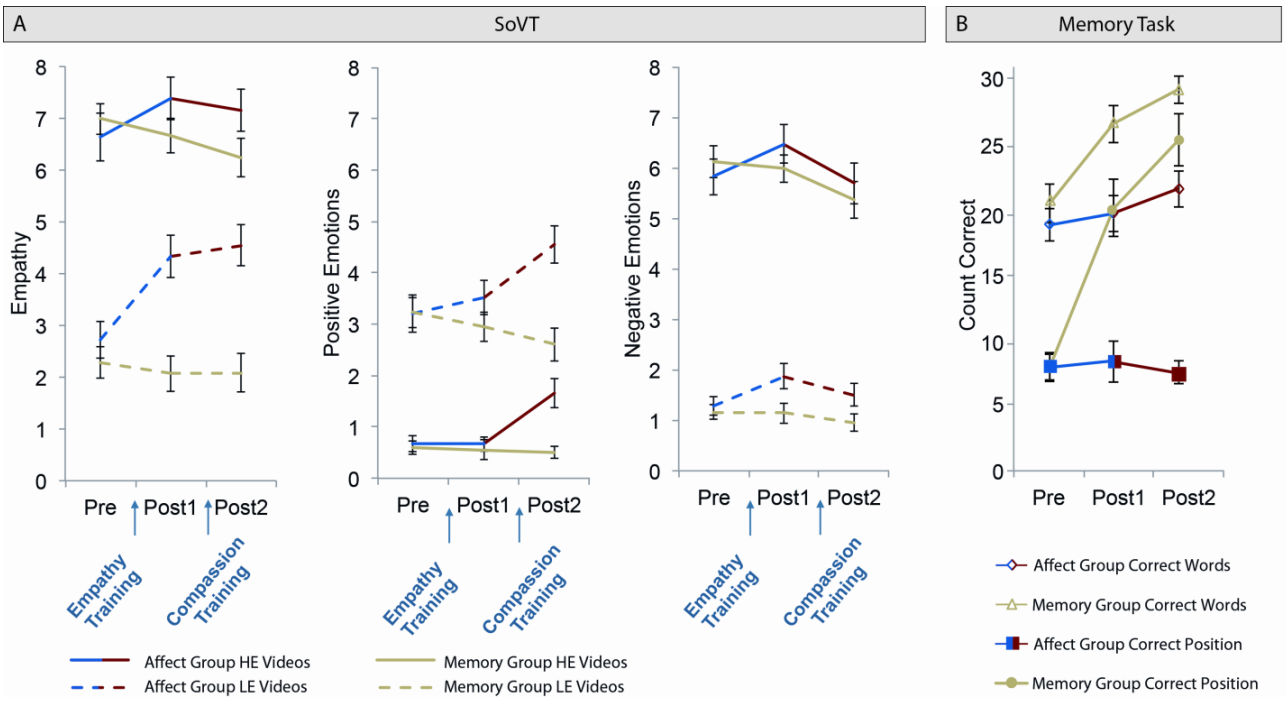


Figure 4.2. (A) Self-reported empathy and negative affect significantly increased after empathy training. Positive affect only increased after compassion training. (B) Memory, but not affect training, improved the number of correctly remembered words and the number of words remembered in the correct position. Error bars indicate standard error of mean; LE, low emotion; HE, high emotion videos.

Functional imaging changes. To examine functional neural plasticity induced by training empathy and compassion, we analyzed imaging data using event-related statistics on the whole brain (p < .05, FWE corrected using cluster level thresholds; Friston, Worsley, Frackowiak, Mazziotta, & Evans, 1994). On the 2nd level, we performed repeated measures ANOVAs with the within subject factors change (2 levels: Pre to Post1, Δ1, and Post1 to Post 2, Δ2) and video type (LE and HE) and the between subject factor training group (affect and memory) (Table S4.2). Age and depression scores (Beck et al., 1996) were included as covariates. Activations in cingulate cortex were classified and labeled based on Vogt (Vogt, 2005). Paralleling behavioral findings, in which similar changes occurred for HE and LE videos, no significant triple interaction

for video type x time x group was observed. Empathy training, but not memory training [Empathy (Δ1 LE & Δ1 HE) > Memory (Δ1 LE & Δ1 HE)], increased activations in insula, temporal gyrus, operculum, posterior putamen, pallidum, thalamus, and head of caudate. In response to HE videos (Empathy Δ1 HE > Memory Δ1 HE) empathy training augmented activations in brain areas spanning insula, aMCC, dorsolateral prefrontal cortex, posterior putamen, pallidum, and head of caudate (Figure 4.3). The same contrast for LE videos revealed overlapping changes in right anterior and middle insula, operculum, temporal gyrus, and head of caudate. Notably, activation changes for HE videos after empathy training overlapped with meta-analytic findings on empathy for pain (Lamm et al., 2011) in aMCC and left AI (Figure 4.3A). Conversely, compassion training, but not a second day of memory training [Compassion (Δ2 LE & Δ2 HE) > Memory (Δ2 LE & Δ2 HE)], increased activations in mOFC, pregenual anterior cingulate cortex (pACC), inferior frontal gyrus (IFG), and ventral striatum. Consistently, in response to HE videos, compassion training augmented activity in mOFC, pACC, ventral striatum, and right middle frontal gyrus. This overlapped with changes induced by compassion training in response to LE videos in pACC and right IFG. To visualize the change in parameter estimates for each factor, mean activations across all voxels in each cluster were extracted (Figure 4.3B). None of the contrasts comparing changes induced by memory training with changes induced by affect training revealed significant activations. To test for any overlap between regions showing functional plasticity after empathy and compassion training, we conducted conjunction analyses of equivalent empathy- and compassion-related contrasts (Table S4.2). None of the conjunctions were significant, underlining that patterns of induced functional plasticity after empathy and compassion training were distinct and non-overlapping.

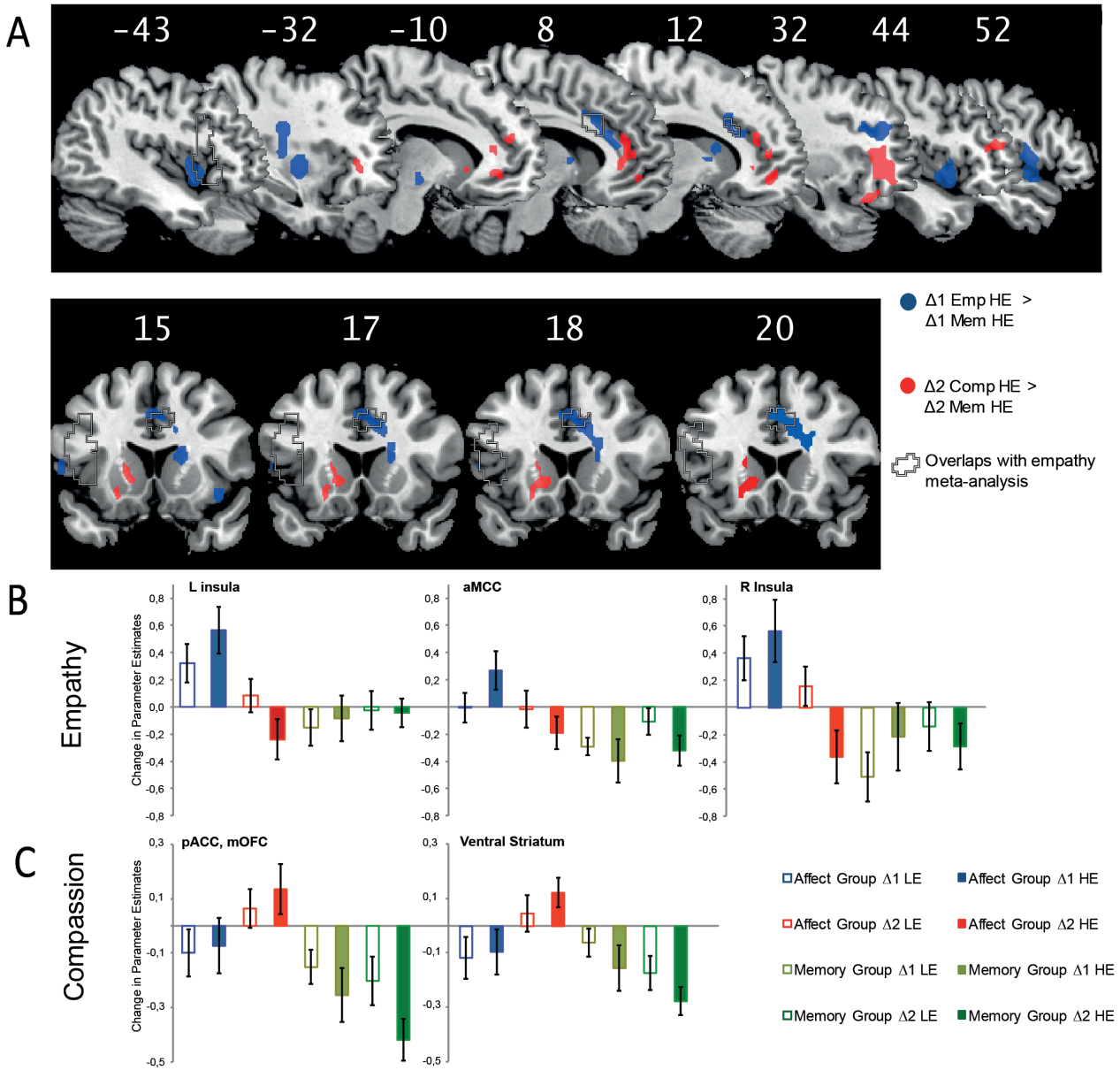


Figure 4.3. (*A*) Functional neural changes related to empathy (blue) and compassion training (red) in comparison with the memory control group ($p < .05$, FWE corrected). Regions in which changes related to empathy training overlap with a recent empathy for pain meta-analysis (Lamm et al., 2011) are indicated by dashed lines. (*B*) Bar charts of changes in parameter estimates in the areas related to empathy training (Empathy Δ1 HE > Memory Δ1 HE). (*C*) Bar charts of changes in parameter estimates of the areas related to compassion training (Compassion Δ2 HE > Memory Δ2 HE). The values represent the mean activation of all voxels in one cluster; error bars depict standard error of mean.

Parametric analyses in the affect group revealed that the increase in empathy ratings after empathy training (Δ1) correlated with the increase in aMCC activity ($p < .05$, FWE corrected, Table S4.3). Activation changes in right AI were also parametrically modulated by increases in negative affect and empathy ratings between Pre to Post1, albeit at an uncorrected threshold of $p < .001$. No significant correlations were found for changes in positive affect ratings after empathy training. Parametric analyses on the changes in subjective ratings and brain activity after compassion training (Δ2) revealed that activity changes in the left supramarginal gyrus ($p < .05$, FWE corrected) were linearly modulated by changes in negative affect ratings. No significant effects were found for changes in empathy ratings or positive affect ratings after compassion training. Practice duration did not parametrically modulate neural changes. The intervention was probably too short to reveal a robust impact of inter-individual practice differences on neural and experiential changes.

In summary, observed changes in brain activation after empathy and compassion training revealed distinct patterns of functional brain plasticity. The effects of empathy training overlapped with previous peak activations in AI and aMCC as identified in a meta-analysis on cross-sectional empathy for pain studies (Lamm et al., 2011). Subsequent compassion training induced activations in a non-overlapping network spanning mOFC, pregenual ACC and ventral striatum.

DISCUSSION

The goal of this short-term affective intervention study with an active memory control group was to dissociate empathy and compassion and to investigate related plasticity on the neural and experiential level. We hypothesized that although these two socio-affective and motivational states may be related, they may have important differential signatures and consequences. Thus,

we anticipated that empathizing with the suffering of others might be associated with negative states, distress and activations in brain networks playing a crucial role in empathy for pain. Conversely, compassion should be accompanied by positive feelings of warmth and concern for the other and increased activations in brain networks related to reward and affiliation.

Indeed, we found evidence for different patterns of emotional experiences and neural plasticity associated with the sequential training of these two social emotions within the same participants: A short-term training in empathy increased empathic responses and negative affect in response to others' distress. In addition, watching others' suffering after empathy training was associated with activations in a network spanning insula, aMCC, temporal gyrus, DLPFC, operculum and parts of basal ganglia. These results align with and extend previous cross-sectional meta-analytic findings on a crucial role of insula and aMCC in empathy for pain (Fan et al., 2011; Lamm et al., 2011) as well as their involvement in self-experienced pain, and negative affect in general (Beckmann et al., 2009; Shackman et al., 2011; Lamm et al., 2011).

Importantly, compassion training reversed these effects: it decreased negative affect back to baseline levels and increased positive affect. On the neural level, compassion training increased brain activations in mOFC, pregenual ACC, and striatum – a network previously associated with positive affect (Kringelbach and Berridge, 2009), affiliation (Strathearn et al., 2009) and reward (Haber and Knutson, 2010). Interestingly, this distinction is paralleled by recent neuroscientific evidence which indicates that social connectedness is typically associated with activations in brain regions that comprise ventromedial prefrontal cortex and ventral stiatum, while social disconnection is rather associated with activations in AI and dorsal ACC (for review, see Eisenberger and Cole, 2012).

The analyses of subjective ratings revealed that empathy training led to an increase in empathy and negative affect in response to both, LE and HE videos. This suggests that training empathy not only induced a stronger sharing of painful and distressing experiences, but even increased the susceptibility to feel negative affect in response to everyday life situations.

Importantly, compassion training counteracted this effect: it increased positive affect and decreased negative affect back to baseline levels. Remarkably, the increase in positive affect occurred even though participants were still exposed to equally distressing video material. This finding adds to the observation of a previous study in which a similar compassion and loving kindness training increased general levels of positive affect in daily life (Fredrickson et al., 2008). It is also in line with previously observed experiential and neural effects after a short-term compassion training (Klimecki et al., 2012). Taken together, this suggests that the generation of compassion in response to distressing situations is distinct from other emotion regulation strategies, like suppression or reappraisal, which involve an active down-regulation of negative affect (Gross, 2002). Thus, the generation of compassion focuses on strengthening positive affect, while not ignoring the presence of suffering or changing the negative reality. Future studies may formally compare compassion with existing emotion regulation strategies. As compassion does not rely on the temporal denial of the negative nature of events, one hypothesis would be that compassion training would abolish rebound effects, as observed in the amygdala after effortful emotion regulation (Walter et al., 2009). Furthermore, acknowledging the negative experience of others rather than suppressing it may be a crucial prerequisite for the development of prosocial motivation and helping behavior. Accordingly, it has recently been shown that the frequency of helping behavior can indeed be increased with a similar short-term compassion training (Leiberg et al., 2011).

On the neural level, we obtained evidence that short-term empathy training, but not memory training, induced functional plasticity in a network spanning insula, aMCC, temporal gyrus, operculum, DLPFC, posterior putamen, pallidum and head of caudate. The observed activation increases in DLPFC and middle temporal gyrus align with previous findings on emotion regulation (Kalisch, 2009), cognitive control (Beckmann et al., 2009; Mansouri et al., 2009; Shackman et al., 2011) and pain processing (Beckmann et al., 2009; Shackman et al., 201).

Importantly, and as illustrated in Figure 4.3A, activations in AI and aMCC were concordant with peak activations identified in a meta-analysis performed over more than 30 cross-sectional studies on empathy for pain (Lamm et al., 2011). Moreover, activation in AI and aMCC have repeatedly been observed to covary with negative affect ratings, both during self-experienced pain and when the unpleasantness of observing others suffering (Akitsuki and Decety, 2009; Cheng et al., 2007; Jackson et al., 2005; Lamm et al., 2007; Saarela et al., 2007; Singer et al., 2008). Finally, we previously observed that activations in AI and aMCC are parametrically modulated by individual differences in empathic experiences for distressing videos in the SoVT (Klimecki et al., 2012). However, the activation changes stemming from empathy training were not limited to AI, but instead spanned the entire insular cortex. This accords with a key role of insular cortex in integrating interoceptive information (Craig, 2009; Lamm and Singer, 2010).

In contrast to empathy training, cultivating feelings of kindness, warmth, and concern induced non-overlapping brain changes in mOFC, pACC, and striatum. These findings extend previous functional imaging findings on compassion in cross-sectional studies (Beauregard et al., 2009; Kim et al., 2009) and one short-term intervention study from our group (Klimecki et al., 2012). For example, viewing sad facial expressions with a compassionate stance was observed to activate ventral striatum and VTA/SN (Kim et al., 2009). In addition, activations in the head of caudate and VTA occured when participants applied unconditional love towards pictures of intellectually disabled individuals (Beauregard et al., 2009). Finally, the present results mirror our previous findings on the effects of short-term compassion training (Klimecki et al., 2012) in a network involving mOFC, striatum and VTS/SN. In general, mOFC, pACC, and ventral striatum activations have been shown to be centrally implicated in reward processing (Haber and Knutson, 2010) as well as in the experience of pleasure and positive affect (Kringelbach and Berridge, 2009). In addition to this convergence with previous neuroimaging findings on positive affect and reward, activations in prefrontal cortex and ventral striatum have been related more

specifically to maternal affiliation (Strathearn et al., 2009), as well as to maternal and romantic love (Bartels and Zeki, 2004). Similarly, studies in rodents and other mammals suggest that the formation of affiliative memories relies on a circuitry comprising mOFC, ventral striatum, and ventral tegmental area (Depue and Morrone-Strupinsky, 2005). Furthermore, animal models distinguish between different affective and motivational systems such as panic and care systems that rely on distinct brain networks and neurotransmitter systems (Panksepp, 2011). In line with this notion, our results suggest that empathy and compassion indeed rely on antagonistic affective systems and that even short-term training of compassion has the potential to counteract empathic distress.

The observed increases in brain activation after compassion and empathy training also differed with respect to their location in the cingulate cortex. Empathy training led to an increase of activation in aMCC. A recent meta-analysis of 939 studies (Shackman et al., 2011) found that aMCC is crucial for processing negative affect, pain and cognitive control. Converging results were provided by a different meta-analysis (Beckmann et al., 2009) which reported that aMCC is implicated in processing pain, conflict monitoring and error detection. In addition, this part of cingulate cortex was found to be highly connected to dorsal prefrontal regions (Beckmann et al., 2009). In keeping with this structural connectivity, training empathy in the current study increased activations in both aMCC and DLPFC. The comparison of cingulate cortex locations from fMRI studies on reward processing revealed a more anterior activation (Beckmann et al., 2009), which converges with the present observation of pACC involvement in compassion. Consistent with this notion, this part of the cingulate cortex was shown to be highly connected with ventral striatum and OFC (Beckmann et al., 2009).

In summary, the present findings reveal that already short-term affective intervention programs can induce reliable experiential and neural plasticity. More importantly, we could show that training two seemingly similar social emotions altered brain activation in non-overlapping neural

networks and changed affective responses of opposing valence. Whereas empathy training increased negative affect and activation in associated brain circuits, compassion training reversed these effects by strengthening positive affect and activation in networks associated to affiliation and reward. Compassion may therefore represent a very potent strategy for preventing burnout. In light of high prevalence rates of burnout and stress-related diseases in Western societies, we anticipate that the present findings will inform other intervention studies on the plasticity of adaptive social emotions. As the present study only focused on females, future studies are needed to address whether the observed training effects can also be generalized to the male population. In the long run, the gained insights will hopefully help to design new training programs aimed at increasing resilience and coping strategies in many domains, including health care, educational settings, and high-stress environments in general.

Acknowledgements

We thank the participants, the people who assisted in conducting the experiments, Claus Lamm for advice and Stephan Liebig for help with the figures. Our gratitude especially goes to the teachers Catherine Felder, Annette Rentz-Lühning, Jotika Hermsen, and Ariya Ñani. This work was supported by grants from the University of Zurich, the Swiss National Science Foundation and the European Research Council (205557).

Author Contribution Statement

T.S., O.M.K., S.L. and M.R. designed research. O.M.K. and S.L. performed research. O.M.K. analyzed the data. O.M.K., T.S., and S.L. wrote the paper.

Supplementary Material

METHODS

Participants. Inclusion criteria for this study were: Toronto Alexithymia Scale (Bagby et al., 1994) < 60, Beck's Depression Inventory (Beck et al., 1996) < 18, no contraindication for fMRI, and no current psychiatric illness as determined by a Structured Clinical Interview for DSM Disorders (SCID (Wittchen et al., 1997)) via phone. Empathy was assessed through the Interpersonal Reactivity Index (Davis, 1983), compassion by means of the Compassionate Love Scale (Sprecher and Fehr, 2005). With regard to highest completed education, in the affect group, 2 participants completed secondary school or apprenticeship (2 in the memory control group), 12 completed high school (21 in the memory control group) and 11 completed university or higher education (5 in the memory control group). The five persons who dropped out in the memory control group had higher scores on the Beck's Depression inventory than those who completed the study ($t(31) = 2.31$, $p < .05$, dropouts: mean = 7.4, SD = 5.6; completers: mean = 3.11, SD = 3.49). No other selective dropouts were observed in the memory control group and no selective dropouts occurred in the affect group. To account for selective dropout and age differences between both groups (Table S4.1), we included age and depression scores as covariates in the analyses. In one participant of the affect training group, subjective ratings at Post1 from the first run were overwritten by the second run due to an irreversible program error. Thus, subjective ratings at Post1 were available for 12 instead of 24 videos for this person. While all main behavioral and imaging analyses with the SoVT were computed with the final sample of 53 participants ($n = 25$ in affective group, $n = 28$ in memory group), the sample size for the behavioral analyses of the memory task (Figure 4.2) consisted of 51 participants (affect training group $n = 24$; memory training group $n = 27$), due to two missing Pre-training measurements of the memory task (one in each group).

Training Courses. Training days in both groups consisted of a structured one-day course (6h) led by a teacher with over ten years of teaching experience in the respective field. Each training started with a brief theoretical introduction (15-30 min) to the training method. Following this introduction, participants in both groups practiced mental imagery in silent periods of 15 to 30 min. Feedback sessions were held in between training periods so that questions pertaining to the practice could be clarified. At mid-day a 45 min lunch break was held. Thus, the structure of the cognitive and affect training days was closely matched. In terms of practice method, both the cultivation of emotions and the memory training relied on mental visualization practices. The main difference between the training regimes was that while the affect training group cultivated their social emotions, the memory control group learned a cognitive skill. Training days in both groups were held five days apart. Pre-training measurement (Pre) was acquired 2 to 13 days before the first training (8.81 ± 2.44 days, mean ± SD). Measurements after the first training (Post1) were acquired one to four days after empathy/memory training (2.74 ± 0.96 days, mean ± SD). Measurements after the second training were taken one to six days after compassion/memory training (3.23 ± 1.2 days, mean ± SD).

Empathy Training. Empathy training aimed at enhancing the ability to empathize with the suffering of other persons. After a theoretical introduction, the teacher guided participants to visualize their own difficult experiences in the past and to get in touch with their own past suffering. Sentences like, “I am touched by my suffering” and “I am in contact with my pain” aided the practice. In the next session, participants were guided to visualize a close person who was not well and to empathize with this person’s suffering. To aid the practice, participants could use internally generated sentences like, “I share your suffering”, or “I see your pain”. This affective resonance with the suffering of others was then sequentially extended towards a neutral person, towards strangers and towards all human beings. To prevent exhaustion and to

enable a continuous practice, participants were given the option of modulating the intensity with which they immersed themselves in empathy for suffering.

Compassion Training. The employed compassion-enhancing technique focused on the practice of loving kindness and is based on an Eastern contemplative tradition (Salzberg, 2002; Singer and Bolz, in press). Previous studies used this method to investigate the effects of compassion training on emotional well-being (Fredrickson et al., 2008), immune function (Pace et al., 2008), prosocial behavior (Leiberg et al., 2011) and neural function (Klimecki et al., 2012; Weng et al., in press). In addition, several cross-sectional studies investigated the effects of long-term compassion practice onto neural and physiological responses to emotional stimuli (Lutz et al., 2008; 2009). The practice of loving kindness or compassion aims at fostering an attitude of friendliness towards oneself and others. After a theoretical introduction, participants were guided to visualize their own past suffering and to relate to it with feelings of warmth and care. Sentences like, “May I be sheltered by compassion” or “May I be safe” aided this practice. This caring affective experience was then sequentially extended towards a close person, a suffering person, a neutral person, and finally towards strangers and human beings in general. The ultimate goal of this training was to foster the capacity to experience care and warmth, even when faced with one’s own suffering or the suffering of others. Both the empathy and the compassion training practices were carried out in silence and participants practiced either in an upright, seated posture or while walking. Guided training periods typically lasted between 15 and 30 min.

Memory Training. To achieve comparable conditions, the active control group attended two one-day courses (6h each) of memory training. Participants were instructed in the Method of Loci, a technique to memorize items in an ordered sequence (Bower, 1970). The Method of Loci relies on memorizing item lists by creating vivid mental images that associated each item with a location on a pre-defined route. For instance, in order to associate the word “penguin” with the

main station, one would imagine a huge penguin sitting on top of the station. Item lists are recalled by remembering the previously created mental images when sequentially revisiting the locations. Notably, the memory training also relied on silent practice periods which lasted 15 to 30 min. These periods of silent visualizations were interspersed with feedback sessions.

More specifically, during the first training day, participants started out by learning to memorize a sequence of 10 self-chosen locations. The morning session was used to memorize this sequence and to learn how to associate word lists with this sequence by constructing vivid mental images. In the afternoon session, participants learned a fixed sequence of 34 locations, which formed a route through Zurich. They memorized this route in three steps and learned to associate words with each of the three sub-routes. The second training day started with a recapitulation of the learned content from the first training day. After the lunch break, participants learned to associate more abstract terms with the given locations. This was practiced in three sessions. Finally, participants completed several sessions in which they memorized lists of 34 items by means of the Zurich route.

In between measurements, participants practiced the trained method during 1h evening classes, which took place until the last measurement day. Participants who could not join these evening sessions were encouraged to practice at home. Furthermore, participants were encouraged to practice throughout their daily lives and, indeed, several reported engaging in the trained methods, for instance while using public transport or at work. To get an impression of practice commitment, we asked participants to record the duration of their daily practice on a paper and pencil sheet. All participants in the affect training group returned their questionnaires, indicating a mean practice duration of 98.6 min (SD = 60.91) after empathy training and a mean practice duration of 65.72 min (SD = 53.35) after compassion training. Paired sample *t*-tests revealed that participants practiced longer after empathy than after compassion training ($t = 2.38$, $p < .05$). In contrast to the complete data return in the affect group, in the memory control group more

than a quarter of participants (8 out of 28) did not provide this information. Due to this unreliable return, we did not analyse practice reports in this group. To prevent such unreliable return in the future, it would probably be beneficial to rely on computerized measures.

Behavioral data analysis. Subjective ratings of empathy, positive affect and negative affect in response to the SoVT and recollection in the memory task were analyzed using repeated measures multivariate analyses of variance (MANOVA) (Pillai's trace statistics). Age and depressive symptoms (Beck et al., 1996) were entered as covariates to model age differences and effects due to selective dropout in the memory control group. These analyses were followed-up by ANOVAs and *t*-tests. In addition, we assessed Pearson`s correlations between practice duration (between Training1 and Post1 measurement and between Training2 and Post2 measurement) and the corresponding changes in subjective ratings in response to the SoVT.

fMRI data acquisition and analyses. A 3T Philips Achieva scanner with an eight channel sense head coil was used for acquiring blood oxygenation level dependent (BOLD) contrast with a T2*-weighted single-shot echo-planar imaging (EPI) pulse sequence (TR = 1.5s, TE = 30ms, flip angle = 90°, FoV = 240 x 228 mm^2, slice gap = 1 mm). Twenty-nine transverse slices (voxel size = 3 x 3 x 3 mm^3) were obtained in an ascending order using an oblique rotation of 20° relative to the anterior commissure-posterior commissure line. The first five volumes were discarded to allow for signal equilibration. SPM8 software (Wellcome Trust Centre for Neuroimaging, London) on Matlab 7.10 (MathWorks, Natick, MA) was used for fMRI data analysis. We conducted standard preprocessing, including realignment, normalization of each participant's functional data to the Montreal Neurological Institute (MNI) EPI template and smoothing with a Gaussian kernel of 6-mm full-width at half-maximum (FWHM).

General linear models accounting for low frequency drifts (cut-off period 128s) and temporal autocorrelations (AR(1)) were estimated for each participant. For the analysis of training-related changes (n_{affect} = 25, n_{memory} = 28), we modeled the regressors LE videos, HE videos and ratings for each session and time point (Pre, Post1, and Post2). Realignment parameters were entered in the design matrix to correct for residual effects of head motion. Regressors were convolved with a hemodynamic response function. The following linear contrasts were computed on the 1st-level for the analysis of training-related changes: (1) Post1 LE videos > Pre LE videos (Δ1 LE videos), (2) Post1 HE videos > Pre HE videos (Δ1 HE videos), (3) Post2 LE videos > Post1 LE videos (Δ2 LE videos), (4) Post2 HE videos > Post1 HE videos (Δ2 HE videos). Smoothed (6 mm FWHM) 1st-level contrast images were entered into 2nd-level random effects group analyses. Training-related changes in BOLD responses were assessed in repeated measures ANOVAs with the within subject factors change (2 levels: Δ1 and Δ2) and video type (2 levels: LE and HE videos) and the between subject factor training group (2 levels: affect training and memory training). To account for the between group differences in age and depression (Beck et al., 1996), age and BDI scores were included as covariates. Corrections for multiple comparisons were carried out at the cluster level using Gaussian random field theory (Friston et al., 1994) [minimum voxel *z*-score ≥ 3.28; cluster significance, $p < .05$, FWE corrected].

To probe the degree to which the changes induced by empathy training as compared to memory training overlap with changes induced by compassion training as compared to memory training, we carried out the following four conjunction analyses: i) [Empathy (Δ1 HE & Δ1 LE) > Memory (Δ1 HE & Δ1 LE)] AND [Compassion (Δ2 HE & Δ2 LE) > Memory (Δ2 HE & Δ2 LE)], ii) [Empathy Δ1 HE > Memory Δ1 HE] AND [Compassion Δ2 HE > Memory Δ2 HE], iii) [Empathy Δ1 LE > Memory Δ1 LE] AND [Compassion Δ2 LE > Memory Δ2 LE], iv) [Empathy (Δ1 HE - Δ1 LE) > Memory (Δ1 HE - Δ1 LE)] AND [Compassion (Δ2 HE - Δ2 LE) > Memory (Δ2 HE - Δ2 LE)].

In order to assess consistency with previous cross-sectional meta-analytic results, we overlaid the observed training-induced changes with a recent image-based meta-analysis of cross-sectional empathy for pain studies (Lamm et al., 2011) (extent threshold of k = 15). As meta-analyses in the domain of compassion research are still lacking, we restrict ourselves to comparing the current results with previous findings in the discussion.

Parametric modulation of changes in ratings were analyzed for each participant by setting up 1st-level models with the regressors video, trial-by-trial rating and rating phase for each session and time point (Pre, Post1, and Post2). This was done separately for ratings of empathy, positive affect, and negative affect. Models accounted for low frequency drifts (cut-off period 128s) and temporal autocorrelations (AR(1)). To further correct for residual effects of head motion, realignment parameters were entered in the design matrix. Regressors were convolved with a hemodynamic response function. To examine how neural activity was parametrically modulated by changes in subjective ratings in the affect training group, we set up first-level models with the regressors video, trial-by-trial rating, and rating phase separately for ratings of empathy, positive, and negative affect. Smoothed (6mm FWHM) 1st-level contrast images of the parametric modulation differences Post1 > Pre and (Δ1) and Post2 > Post1 (Δ2) were entered into 2nd-level one-sample *t*-tests. In order to test whether self-reported practice duration parametrically modulated neural changes in the affect group, we set up additional one-sample *t*-tests in which duration was included as a covariate.

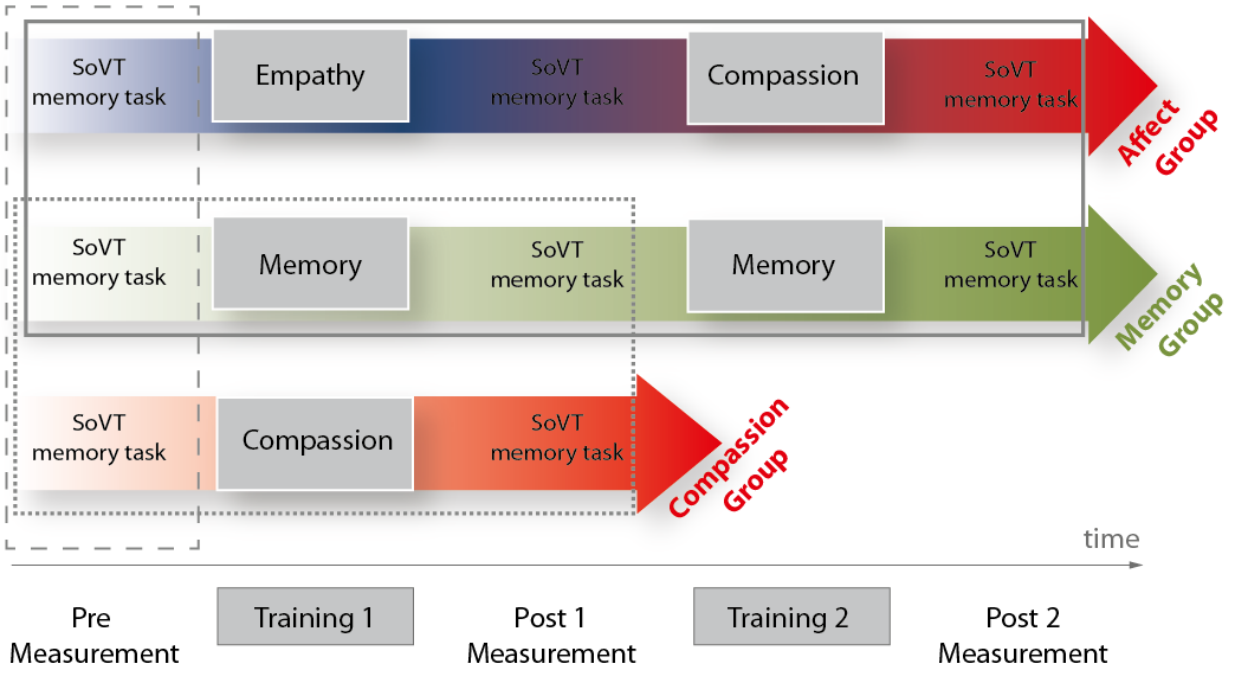


Figure S4.1. Relation to previous studies. Participants whose results are reported here were part of a larger project on affective plasticity (ERC grant) from which two data sets have already been published (Klimecki et al., 2012). The dashed lines indicate that Pre Measurement self-report and neuroimaging data from the SoVT were used for original validation purpose of the newly developed task (Klimecki et al., 2012) (total *N* = 94). The dotted lines show that longitudinal training data from the SoVT and the memory task were used to investigate the effects of compassion training (*n* = 28) in comparison with a matched memory control group (Klimecki et al., 2012) (*n* = 30). The current study investigated changes induced through training empathy and compassion in a new independent affect group (marked by a solid line). To this end, data from the affect intervention group (*n* = 25) were compared to the memory group (*n* = 28) at three time points. In spite of the use of a subset of neuroimaging, behavioral, and self-report data from the memory group in several analyses, the affect intervention group against which these data were compared was always different. This assures that all analyses reported are unique and never redundant. SoVT, Socio-affective Video Task.

Table S4.1

Sample characteristics for participants in the affect and the memory training group

Variable	**Time point**	**Affect Group (*n* = 30; *n* = 25)**	**Memory Group (*n* = 33; *n* = 28)**	**Group Difference, *t*-Test**
Age	pre-training	25.93 (4.27)	22.79 (3.77)	*t(61)* = 3.1, *p* = .003
	completion	25.88 (4.32)	22.89 (4.02)	*t(51)* = 2.61, *p* = .01
Alexithymia[1]	pre-training	40.87 (8.78)	40.58 (7.22)	*t(61)* = 0.14, *p* = .89, n.s.
	completion	41 (8.96)	39.61 (6.57)	*t(51)* = 0.65, *p* = .52, n.s.
Depression[2]	pre-training	5.9 (5.03)	3.76 (4.08)	*t(61)* = 1.87, *p* = .07, n.s.
	completion	6.04 (4.89)	3.11 (3.49)	*t(51)* = 2.54, *p* = .01
Empathy[3]	pre-training	67.8 (9.46)	66.52 (8.92)	*t(61)* = 0.56, *p* = .58, n.s.
	completion	68.68 (8.65)	65.93 (9.44)	*t(51)* = 1.10, *p* = .28, n.s.
Compassionate Love[4]	pre-training	87.1 (21.85)	89.52 (16.86)	*t(61)* = 0.49, *p* = .62, n.s.
	completion	86.28 (19.38)	89 (17.27)	*t(51)* = 0.54, *p* = .59, n.s.

Note. Table depicts means, standard deviations are shown in brackets. [1]Toronto Alexithymia Scale (TAS; Bagby et al., 1994); [2]Beck's Depression Inventory (BDI; Beck et al., 1996); [3]Interpersonal Reactivity Index (IRI (Davis, 1983)); [4]Compassionate Love Scale (CLS (Sprecher and Fehr, 2005)).

Table S4.2

Coordinates of BOLD activation changes differentially related to empathy or compassion training

Anatomical Areas	Cluster size	MNI coordinates (mm)			Z values
		x	y	z	
Empathy (Δ1 HE & Δ1 LE) > Memory (Δ1 HE & Δ1 LE)					
Insula (MI, PI), posterior putamen and pallidum, thalamus, head of caudate	1356	-30	-4	0	4.58
Insula (AI, MI), rolandic operculum, superior temporal gyrus, posterior putamen	881	52	4	-4	4.24
Empathy Δ1 HE > Memory Δ1 HE					
Insula (AI, MI, PI), operculum, posterior putamen, and pallidum	1276	-32	-4	0	4.21
aMCC, head of caudate, MFG (DLPFC)	921	32	38	24	4.07
Insula (AI, MI), rolandic operculum, superior temporal gyrus, posterior putamen	650	52	2	-6	3.98
Empathy Δ1 LE > Memory Δ1 LE					
Insula (AI, MI), rolandic operculum, IFG (pars opercularis), superior temporal gyrus	685	52	0	8	4.02
Head of caudate, thalamus	449	6	0	4	3.46
Empathy (Δ1 HE - Δ1 LE) > Memory (Δ1 HE - Δ1 LE)					
n.s.					
Compassion (Δ2 HE & Δ2 LE) > Memory (Δ2 HE & Δ2 LE)					
mOFC, MFG, pACC, IFG (pars orbitalis)	1605	34	40	-2	4.26
pACC, ventral striatum, IFG (pars orbitalis)	768	-12	38	12	3.7
Compassion Δ2 HE > Memory Δ2 HE					
pACC, mOFC, MFG, IFG (pars orbitalis, pars triangularis)	1598	-10	38	-6	4.12
ventral striatum, pACC, IFG (pars orbitalis)	680	-18	44	10	4.08
Compassion Δ2 LE > Memory Δ2 LE					

pACC, IFG (pars triangularis)	548	16	40	24	3.51
IFG (pars opercularis), dorsal AI, dorsal putamen, and pallidum	498	38	12	22	3.81

Compassion (Δ2 HE - Δ2 LE) > Memory (Δ2 HE - Δ2 LE)

n.s.

Memory (Δ1 HE & Δ1 LE) > Empathy (Δ1 HE & Δ1 LE)

n.s.

Memory Δ1 HE > Empathy Δ1 HE

n.s.

Memory Δ1 LE > Empathy Δ1 LE

n.s.

Memory (Δ1 HE – Δ1 LE) > Empathy (Δ1 HE – Δ1 LE)

n.s.

Memory (Δ2 HE & Δ2 LE) > Compassion (Δ2 HE & Δ2 LE)

n.s.

Memory Δ2 HE > Compassion Δ2 HE

n.s.

Memory Δ2 LE > Compassion Δ2 LE

n.s.

Memory (Δ2 HE - Δ2 LE) > Compassion (Δ2 HE - Δ2 LE)

n.s.

CONJUNCTIONS

[Empathy (Δ1 HE & Δ1 LE) > Memory (Δ1 HE & Δ1 LE)] AND
[Compassion (Δ2 HE & Δ2 LE) > Memory (Δ2 HE & Δ2 LE)]

n.s.

[Empathy Δ1 HE > Memory Δ1 HE] AND
[Compassion Δ2 HE > Memory Δ2 HE]

n.s.

[Empathy Δ1 LE > Memory Δ1 LE] AND

[Compassion Δ2 LE > Memory Δ2 LE]

n.s.

[Empathy (Δ1 HE - Δ1 LE) > Memory (Δ1 HE - Δ1 LE)] AND

[Compassion (Δ2 HE - Δ2 LE) > Memory (Δ2 HE - Δ2 LE)]

n.s.

Note. All values $p < .05$, FWE corrected. L, left; R, right; Δ1, difference Post1 > Pre; Δ2, difference Post2 > Post1; MNI coordinates and *Z* values are given for the grey matter peaks of the respective cluster. pACC, pregenual anterior cingulate cortex; AI, anterior insula; aMCC, anterior medial cingulate cortex; MFG, middle frontal gyrus; DLPFC, dorsolateral prefrontal cortex; FI, frontal insula; IFG, inferior frontal gyrus; MI, middle insula; mOFC, medial orbitofrontal cortex; PI, posterior insula.

Table S4.3

BOLD activation changes parametrically related to changes in self-reported feelings in the affect group (n = 25)

Brain regions	Cluster size	MNI coordinates (mm)			*Z* values
		x	y	z	
Δ1 Empathy Ratings					
Anterior medial cingulate cortex	468	-4	20	24	4.22
Δ1 Negative Affect Ratings					
n.s.					
Δ1 Positive Affect Ratings					
n.s.					
Δ2 Empathy Ratings					
n.s.					
Δ2 Negative Affect Ratings					
Supramarginal gyrus	382	-62	-24	20	4.35
Δ2 Positive Affect Ratings					
n.s.					

Note. All values $p < .05$, FWE corrected. L, left; R, right; Δ1, difference Post1 > Pre; Δ2, difference Post2 > Post1; MNI coordinates and *Z* values are given for the grey matter peaks of the respective cluster.

5. Structural Covariance Networks of the Dorsal Anterior Insula Predict Females' Individual Differences in Empathic Responding

Corresponding publication:

Bernhardt, B.*, Klimecki, O.*, Leiberg, S., Singer T. (2013). *Structural Covariance Networks of the Dorsal Anterior Insula Predict Females' Individual Differences in Empathic Responding.* Cerebral Cortex. doi:10.1093/cercor/bht072 (*both authors contributed equally to this work). By permission of Oxford University Press.

Abstract

Previous functional imaging studies have shown key roles of the dorsal anterior insula (dAI) and anterior midcingulate cortex (aMCC) in empathy for the suffering of others. The current study mapped structural covariance networks of these regions and assessed the relationship between networks and individual differences in empathic responding in 94 females. Individual differences in empathy were assessed through average state measures in response to a video task showing others' suffering, and through questionnaire-based trait measures of empathic concern. Overall, covariance patterns indicated that dAI and aMCC are principal hubs within prefrontal, temporolimbic, and midline structural covariance networks. Importantly, participants with high empathy state ratings showed increased covariance of dAI, but not aMCC, to prefrontal and limbic brain regions. This relationship was specific for empathy and could not be explained by individual differences in negative affect ratings. Regarding questionnaire-based empathic trait measures, we observed a similar, albeit weaker modulation of dAI covariance, confirming the robustness of our findings. Our analysis, thus, provides novel evidence for a specific contribution of frontolimbic structural covariance networks to individual differences in social emotions beyond negative affect.

Previous functional imaging research has advanced our understanding of the neuronal underpinnings of empathy, the ability to share the feelings and sensations of others (de Vignemont & Singer, 2006). In studies localizing functional correlates of empathy for pain, the dorsal anterior insula (dAI) and anterior midcingulate cortex (aMCC), in particular, have been shown to play a pivotal role in first-hand and vicarious experience of pain (Fan et al., 2011; Lamm et al., 2011). Nevertheless, accumulating findings suggest that both regions also participate in more general processing related to negative affect, and not empathy for pain specifically (Craig, 2009; Singer et al., 2009; Corradi-Dell'Acqua et al., 2011; Shackman et al., 2011).

Another line of empathy research focusing on brain connectivity has suggested that empathic inferences about other people's mental states rely on the integration of affective information with other social cognitive processes such as action observation and perspective taking. In fact, previous functional connectivity studies in the domain of empathy have shown that dAI and aMCC interact with other networks involved in social cognition, such as medial prefrontal cortex (PFC) (Zaki et al., 2007) and inferior frontal gyrus of ventrolateral PFC (Jabbi et al., 2008). Whether such interactions between dAI and aMCC and other routes of social cognition are reflected in inter-regional structural brain networks, and thus possibly relate to individual differences in empathic responding, remains unclear. Studies based on tract tracing, the gold standard for connectivity mapping, have indicated a central role of dAI and aMCC within multiple brain networks in frontolimbic, temporal, and also midline regions (Mufson & Mesulam, 1982; Pandya, Van Hoesen, & Mesulam, 1981). However, as these invasive techniques cannot be applied to humans, so far little is known about structural networks relating to empathy in humans.

To close this gap, we mapped structural covariance networks centered on dAI and aMCC and studied how these are modulated by individual differences in empathic responding. Specifically, we used covariance analysis of MRI-based cortical thickness measurements, a method that has previously been used to reliably map structural brain networks in vivo (Lerch et al., 2006; Bullmore & Sporns, 2009; Seeley et al., 2009). Seed-regions in dAI and aMCC were identified based on a recent meta-analysis of functional studies on empathy for pain (Lamm et al., 2011). To investigate the relationship between structural covariance networks and individual differences in empathy, we calculated average empathic state ratings using a newly developed Socio-affective Video Task (SoVT) that depicts people suffering (Klimecki et al., 2012). Employing dynamic social scenes, as in the SoVT, instead of simplified and artificial stimuli

might be advantageous in assessing the different component networks interacting in empathy (Zaki & Ochsner, 2012). A self-report empathy trait questionnaire was also administered (Davis, 1983). As evidence is accumulating that both dAI and aMCC might play a more general role in negative affect processes (Craig, 2009; Singer et al., 2009; Corradi-Dell'Acqua et al., 2011; Shackman et al., 2011), we tested for the specificity of empathy-related covariance modulations by controlling for individual differences in average negative affect state ratings during the SoVT.

Methods

Subjects. We studied 94 healthy, right-handed women aged between 18 and 35 years (mean ± SD = 24.3 ± 4.17 years) from the University of Zurich and the surrounding community. Only females were recruited due to previously reported gender differences in social emotions on the behavioral and neural level (Baron-Cohen & Wheelwright, 2004; Singer et al., 2006). Prior to participation, all volunteers completed an online version of the Beck Depression Inventory (BDI; Beck et al., 1996), and the Toronto Alexithymia Scale (TAS-20; Bagby et al., 1994), the latter as a measure of difficulties in experiencing and expressing feelings. To ensure that our sample population represented a normally functioning female, only volunteers with BDI <18 and TAS-20 <60 and with no contraindication for MRI were included. We, furthermore, conducted a Structured Clinical Interview for DSM-IV disorders (SCID; Wittchen et al., 1997) to screen for Axis-I and psychotic disorders via phone. Volunteers with these psychiatric illnesses were excluded.

The study was approved by the Research Ethics Committee of Zurich ("Kantonale Ethikkommission des Kantons Zürich, Spezialisierte Unterkommission Psychiatrie, Neurologie, Neurochirurgie"; E-25/2008), and was carried out in compliance with the Declaration of Helsinki. All participants gave written informed consent, were paid for their participation and were debriefed after the study was completed.

Measuring inter-individual differences in empathic responding. We quantified individual differences in empathic responses through both, average empathy state ratings during the SoVT (Klimecki et al., 2012) and self-reported trait empathic concern as measured by the Interpersonal Reactivity Index, IRI (Davis, 1983) questionnaire.

a) Empathy state ratings during the SoVT (EMP-SoVT). To assess average empathic state ratings, we used the newly developed Socio-affective Video Task (SoVT) which is described in detail elsewhere (Klimecki et al., 2012). It consists of 3 sets of 24 videos from documentaries or newscasts (duration 10-18 s): half of the videos depict people in distress (high emotion videos), the other half of the videos serve as a high-level baseline and depict people in everyday situations (low emotion videos). All subjects completed the SoVT during a functional MRI experiment; the functional imaging results are described elsewhere (Klimecki et al., 2012). After each video, participants were asked to rate on a Likert scale ranging from 0, none, to 10, intense, how much empathy, negative affect and positive affect they felt while watching the video. To control for each subject's individual baseline, we then calculated the mean empathy rating difference between high and low emotion videos (EMP-SoVT) and used the obtained values as regressors for subsequent statistical analysis. In a similar fashion, we calculated differences for positive affect ratings (POS-SoVT) and negative affect ratings (NEG-SoVT) as additional control covariates.

b) Self-reported questionnaire measure of trait empathic concern (IRI-EC). All 94 subjects completed the IRI using the online survey program Unipark (Globalpark GmbH, Hurth, DE). The IRI is one of the most commonly used self-report questionnaires of dispositional empathy and contains 7 items in each of the 4 categories: empathic concern, perspective taking, personal distress, and fantasy. Each item consists of a statement to which subjects are asked to indicate on a 5-level Likert-scale whether it describes them well (scored with 4 points) or not at all (scored with 0 points). For main statistical analysis, we selected the empathic concern subscale (IRI-EC). This subscale most closely matches our and others' theoretical conception of empathy (Davis, 1983; Singer, 2006; Singer & Lamm, 2009), as it captures feelings of sympathy and concern for unfortunate others. We thus additionally tested whether this subscale relates to structural covariance measures of dAI and aMCC. The other IRI subscales (Fantasy, Personal Distress, and Perspective Taking) were not assessed, as this also would considerably increase the amount of statistical comparisons performed.

MRI Acquisition. High-resolution T_1-weighted images were obtained using a 3-Tesla Philips Achieva whole-body magnetic resonance imaging system equipped with an 8-channel SENSE head coil (Philips Medical Systems, Best, The Netherlands). Images were acquired using a 3D-TFE pulse sequence with a flip angle of 8°. The 301 slices with a voxel size of 1.1 × 1.1 × 0.6 mm^3 were sagittally placed along the anterior–posterior commissure plane. The acquisition matrix ranged over 240 × 240 with a field of view of 250 × 250 mm^2.

MRI-based cortical thickness measurements. We used FreeSurfer software to generate models of the cortical surface and to model cortical thickness from the T_1-weighted images (Version 5.0.0; http://surfer.nmr.mgh.harvard.edu). Previous work has validated FreeSurfer by comparing it with histological analysis (Rosas et al., 2002) and manual measurements (Kuperberg et al., 2003). The processing steps have been described in detail elsewhere (Dale et al., 1999; Fischl et al., 1999; Han et al., 2006). Following surface extraction, sulcal and gyral features across individual subjects were aligned by morphing each subject's brain to an average spherical representation that allows for accurate matching of cortical thickness measurement locations among participants, while minimizing metric distortion. The entire cortex of each participant was visually inspected and segmentation inaccuracies were manually corrected. For whole-brain analysis, thickness data were smoothed on the tessellated surfaces using a 20-mm full-with-at-half-maximum Gaussian kernel prior to statistical analysis. Selecting a surface-based kernel reduces measurement noise but preserves the capacity for anatomical localization, as it respects cortical topological features (Lerch & Evans, 2005).

Seed definition Based on a previously published meta-analysis on empathy for pain (Lamm et al., 2011), we chose the following seed regions for structural correlation analysis: left dAI (MNI *x*/*y*/*z*: −40/22/0), right dAI (39/23/−4), and left aMCC (−2/23/40). These regions have been identified as the most consistently activated areas in a coordinate-based meta-analysis of 32 functional MRI studies that investigated the neural substrates of empathy for pain (Lamm et al., 2011). Indeed, this study of Lamm et al. (2011) showed dAI activations in 28/32 studies and 24/32 previous functional studies, while other regions were activated in <30% of assessed studies.

Thus, while not addressing all regions possibly involved in processes related to empathy (particularly those involved in the interaction between empathy and socially relevant skills, such as cognitive perspective taking or self-other distinction, including the medial prefrontal cortex or temporoparietal junction area [Saxe & Kanwisher, 2003; Frith & Frith, 2006]), we investigated structural networks centered on the most consistent regions. To avoid a bias to the left hemisphere for the midline region aMCC, we also seeded from the right aMCC counterpart (2/23/40), resulting in a total of 4 seed regions for all subsequent analyses. For each meta-analysis coordinate, we identified the closest surface point on the mesh representation at mid-thickness of FreeSurfer's fsaverage7 template in MNI space. In each subject, the surface-

registered and smoothed cortical thickness measure was then used to construct the seed regressor.

Statistical analyses. Analysis was performed using the SurfStat toolbox (Worsley et al., 2009) for Matlab (R2010a, The Mathworks, Natick, MA, USA).

a) Mapping of structural networks by cortical thickness correlations. To map structural covariance networks involved in empathy in vivo, we correlated the cortical thickness of each seed (i.e., dAI and aMCC in left and right hemisphere) with the thickness across all surface points of the entire cortical mantle. These models were constructed by pairwise correlations of the data of all 94 participants. Following previously reported nomenclature (Lerch et al., 2006), the model fitted for the thickness *T* at a surface point *i* was $T_i = \beta_0 + \beta_1 T_{\text{seed}}$

b) Assessment of relationship between covariance strength and empathy state ratings (EMP-SoVT). We fitted linear models that included main effects of seed thickness, EMP-SoVT (i.e., the difference in mean empathy state ratings for high and low emotion videos), and the parametric interaction term between seed thickness and EMP-SoVT. The model fitted for the thickness *T* at a surface point *i* was $T_i = \beta_0 + \beta_1 T_{\text{seed}} + \beta_2 \text{EMP}_{\text{SoVT}} + \beta_3 (\text{EMP}_{\text{SoVT}} \times T_{\text{seed}})$, where × indicates an interaction between terms. While a positive interaction indicates a stronger structural correlation with EMP-SoVT, a negative interaction indicates a weaker link.

We also examined whether the interactions between structural covariance network strength and interindividual differences in EMP-SoVT were specific to empathy, or whether they could also be accounted for by negative affect ratings during the SoVT. We, thus, repeated the above analysis with NEG-SoVT (i.e., the difference in mean negative affect state ratings for high and low emotion videos) as an additional covariate in the statistical model. Moreover, using a post hoc analysis in findings of EMP-SoVT covariance modulation, we specifically assessed whether covariance strength was also modulated by NEG-SoVT.

To probe the robustness of our findings against diffuse interindividual variations in cortical thickness, the above analyses were repeated with additionally controlling for global mean thickness in the statistical model.

c) Assessment of relationship between covariance strength and self-reported trait empathic concern (IRI-EC). An analysis analogous to *"Assessment of Relationship Between Covariance*

Strength and Empathy State Ratings (EMP-SoVT)" was carried out to assess the positive modulation of structural correlation strength by IRI-EC scores. This model included main effects of seed thickness, IRI-EC scores, as well as the parametric interaction term between seed thickness and IRI-EC scores. The model fitted for the thickness *T* at a surface point *i* was $T_i = \beta_0 + \beta_1 T_{\text{seed}} + \beta_2 \text{IRI}_{\text{EC}} + \beta_3 (\text{IRI}_{\text{EC}} \times T_{\text{seed}})$. In regions of significant parametric interactions of seed covariance by the empathy ratings during SoVT (independently obtained from *b*), we assessed the relationship between seed covariance strength and IRI-EC scores.

d) Correction for multiple comparisons. As in previous work (Bernhardt et al., 2008; Bernhardt et al., 2010), findings from our surface-based covariance analysis were controlled using random field theory for nonisotropic images (Worsley et al., 1999). This framework strongly controlled the chance of "ever" reporting a familywise error (FWE) to be < .05. To illustrate trends, surface maps are also shown at uncorrected threshold holds.

Results

Mapping of empathy networks using structural covariance analysis

Patterns of structural correlations of our seed region in left dAI encompassed large portions of lateral and medial PFC (i.e., ventromedial, ventrolateral, dorsolateral and dorsomedial PFC), cingulate (i.e., subgenual and dorsal anterior cingulate, midcingulate, and posterior cingulate), insular, medial, and lateral temporal, together with occipitotemporal, right temporoparietal, and precuneus regions in both hemispheres ($p < .001$, FWE) (Figure 5.1).

Patterns of structural correlations from the seed in right dAI were similar to the patterns seen for the left dAI, but additionally encompassed bilateral temporopolar regions and posterior parietal midline regions ($p < .001$, FWE).

The seed in left aMCC was correlated with other cingulate subregions, ranging from subgenual anterior cingulate to posterior cingulate cortex ($p < .001$, FWE). Correlations extended to precuneus regions, as well as to a large portion of lateral PFC, right insular cortex ($p < .001$, FWE), and lateral temporal regions ($p < .03$, FWE).

Patterns of right aMCC correlations resembled those of left aMCC, but additionally encompassed extensive and bilateral lateral temporal and insular regions ($p < .05$, FWE).

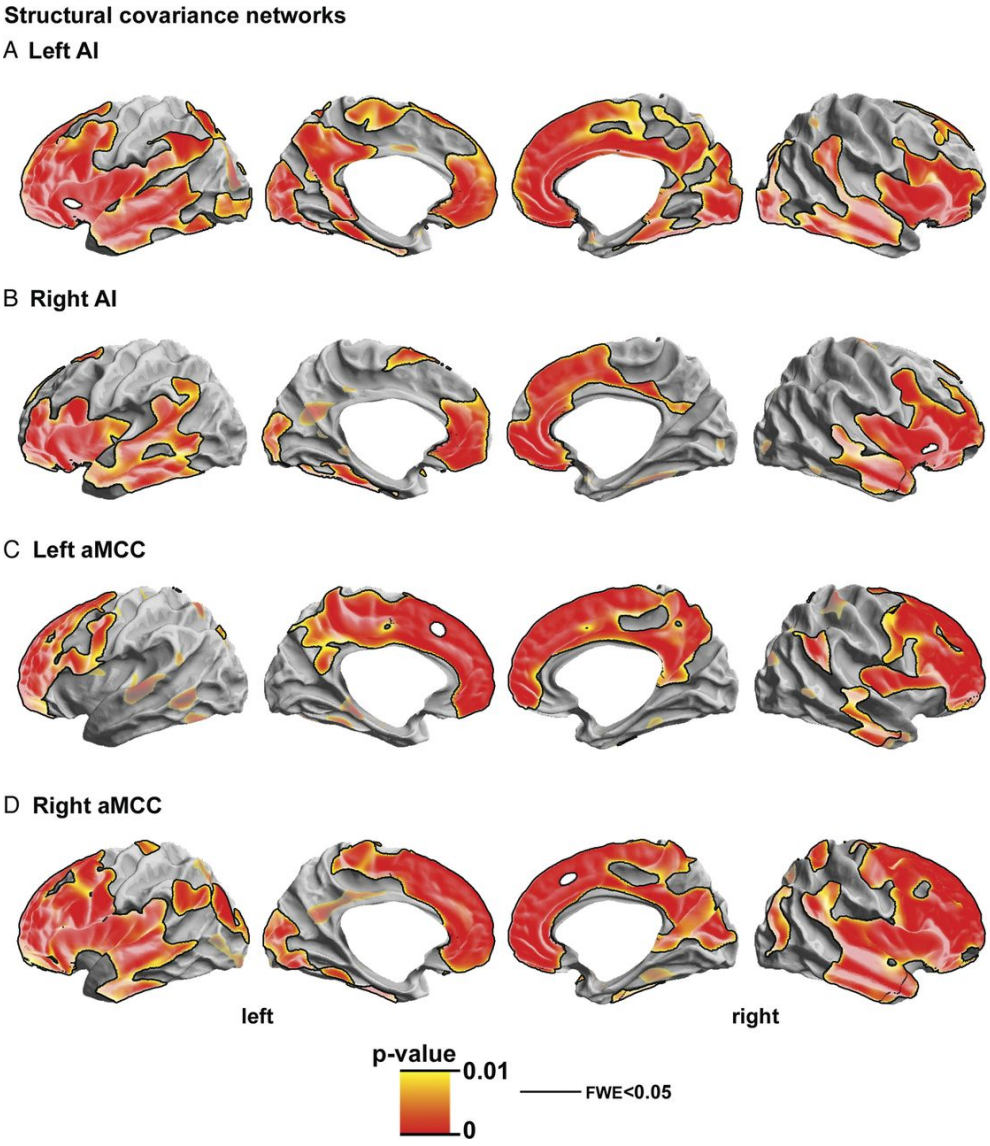


Figure 5.1. Structural covariance analysis seeding from (*A*) left dorsal anterior insula (dAI), (*B*) right dAI, (*C*) left aMCC, and (*D*) right aMCC in 94 females. Seed regions were identified based on a previous meta-analysis of 32 functional MRI studies on empathy for pain (Lamm et al., 2011). Significant correlations between cortical thickness in the seed and a cortical target region across the sample were interpreted as structural networks. To correct for multiple comparisons, significances have been thresholded at $p < .05$, FWE using random field theory for nonisotropic images (cluster threshold = 2.37, extent threshold = 1.75, black outlines). To illustrate trends, findings at $p < .01$, uncorrected (no black outlines, semitransparent) are also shown.

Relationship between structural covariance and EMP-SoVT

We studied the parametric interaction between seed covariance strength and interindividual differences in empathy state ratings during the SoVT (Figure 5.2). This analysis aimed to assess the relationship between structural covariance network configurations and empathic state responses. Please see the Table 5.1 for a detailed overview of the empathy and affect state ratings during the SoVT, as well as our recent functional work (Klimecki et al., 2012). EMP-SoVT ratings were significantly correlated with state negative affect rating differences (NEG-SoVT, r = .51, p < .01, Bonferroni-corrected), but not with state positive affect rating differences (POS-SoVT, r = .03, p > .5).

Structural correlations from our seed in left dAI to target regions in left ventrolateral and anterior PFC were positively modulated by EMP-SoVT ratings (p < .05, FWE, Figure 5.2). This indicated that people with higher EMP-SoVT ratings showed a stronger structural covariance between left dAI and these regions relative to those with lower EMP-SoVT ratings. Considering covariance patterns of the right dAI seed, similar, yet more extended and bilateral effects were observed: EMP-SoVT ratings positively interacted with the covariance strength of the dAI seed to target regions in bilateral anterior and ventrolateral PFC (left: p < .001, FWE; right: p < .03, FWE). In the left hemisphere, this cluster additionally included medial PFC and anterior dorsolateral PFC regions. In the right hemisphere, we observed a separate cluster including mid- and posterior insular regions (p < .02, FWE). Contrary to our expectations, no interactions between aMCC correlations and EMP-SoVT were significant (p > .6, FWE).

Given that EMP-SoVT was significantly correlated to NEG-SoVT (see above), we carried out a series of additional analyses to assess the specificity of the above reported findings for empathy. Analyzing the modulation of dAI covariance strength by NEG-SoVT to each cluster of dAI covariance modulation by EMP-SoVT (see Figure 5.2), we failed to observe any significant modulation of dAI covariance strength by NEG-SoVT (t < 1.4, p > .1). Moreover, statistical model comparison in each cluster between the initial interaction model that assessed parametric modulation of dAI covariance strength by EMP-SoVT and more complex models that additionally included NEG-SoVT as a nuisance regressor failed to indicate that differences in state negative affect ratings (F < 1.24, p > .2) significantly explain additional variance for the dAI covariance modulation by EMP-SoVT to prefrontal and mid-posterior insular regions. These additional analyses suggest that the observed results were indeed specific to empathic responses and could not be accounted for by negative affect alone.

Given that our seed regions displayed marked and widespread structural covariance with other brain networks (see Figure 5.1), we furthermore tested for the robustness of these covariance modulations by EMP-SoVT when controlling for global mean thickness in the same statistical model. Specifically, a post-hoc analysis revealed robust modulations of dAI covariance by EMP-SoVT in all aforementioned clusters even when the model additionally included global mean thickness as a control covariate ($t > 3.0$, see Supplementary Figure 5.1).

Table 5.1

	Min	Max	Mean±SD	Median
EMP-SoVT	-0.08	9.00	4.29±1.84	4.16
High emotion videos	0.67	9.75	6.78±1.93	7.08
Low emotion videos	0.00	5.83	2.49±1.67	2.25
POS-SoVT	-6.25	-0.50	-2.57±1.32	-2.42
High emotion videos	0.00	2.83	0.59±0.68	0.33
Low emotion videos	0.67	6.75	3.16±1.52	2.83
NEG-SoVT	-2.00	8.50	4.78±1.67	4.75
High emotion videos	0.83	9.67	6.04±1.75	5.96
Low emotion videos	0.00	4.00	1.26±0.92	1.00
IRI-EC	12	28	20.42±3.58	20

Note. State ratings consisted of empathy/positive affect/negative affect ratings in response to high and low emotion videos during the Socio-affective Video Task (SoVT). For the regression analysis, the difference between high and low emotion videos were calculated for each individual (EMP-SoVT/POS-SoVT/NEG-SoVT). The empathic concern subscale of the interpersonal reactivity index (IRI-EC), a self-report questionnaire, was used to measure trait empathic concern.

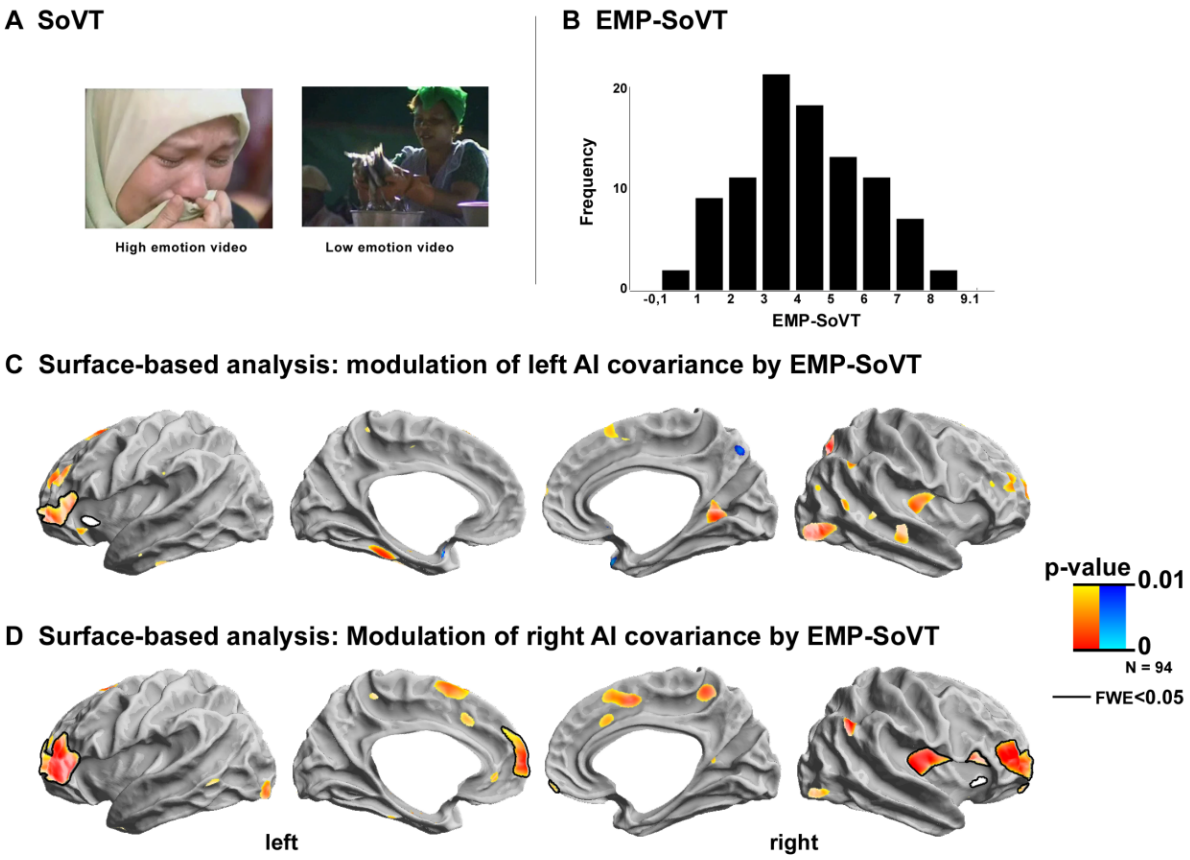


Figure 5.2. Interactions between the degree of structural covariance of dAI seed regions and the empathy state ratings during the socio-affective video task, SoVT (Klimecki et al., 2012). (*A*) Illustration of the SoVT stimuli: Participants viewed 12 high emotion videos (people in distress) and 12 low emotion videos (people in everyday situations), with each video lasting from 10 to 18 s. After each video, participants rated their subjective experience of empathy, positive affect, and negative affect. Every mini-block of 3 high or low emotion videos was followed by a fixation cross, displayed for 10 s. (*B*) Histogram of average EMP-SoVT ratings, which is the difference between empathy ratings during high emotion videos and low emotion videos. (*C* and *D*) Positive interactions between the degree of structural covariance to left and right dAI and EMP-SoVT ratings, indicating higher structural coupling in subjects with higher EMP-SoVT relative to those with lower EMP-SoVT. Please, see Figure 5.1 for details on the statistical thresholding.

Relationship between covariance strength and trait empathic concern (IRI-EC)

To determine whether there was a similar positive modulation of structural covariance networks by self-report trait questionnaire measures of empathic concern, we assessed the interaction between covariance strength and IRI-EC (Figure 5.3). Please see Table 5.1 for additional details on the IRI scores. IRI-EC was positively correlated with EMP-SoVT (r = .23, p < .05), but not with NEG-SoVT (r = .16, p > .1). There was, however, no difference in the strength of correlations (z < 0.6, p > .3).

Analyzing the interaction between dAI and aMCC covariance and IRI-EC scores on a whole brain level, we did not observe any significant positive modulation (p > .14, FWE). In a next step, we carried out region-of-interest analyses of positive IRI-EC modulations on the covariance between dAI and significant clusters from Figure 5.2 (Note that these clusters were generated independently by testing on the interaction between dAI covariance and EMP-SoVT). This analysis aimed at investigating convergences between the covariance modulations based on task- and questionnaire-based empathy measures. Using these regions of interest, we observed that the covariance between the right dAI to right ventrolateral and anterior PFC was positively modulated by IRI-EC (t = 2.1, p < .04, uncorrected, Figure 5.3). These findings indicate a stronger covariance in subjects with high IRI-EC relative to those with low total IRI-EC. No effects were found in the other clusters whose covariance to right and left dAI was modulated by interindividual differences in empathy ratings ($|t|$ < 1.01, p > .1, Figure 5.3).

Virtually identical patterns of findings were observed when assessing covariance modulations by the overall IRI scale instead of IRI-EC. Indeed, overall IRI also selectively modulated covariance strength between right dAI and right ventrolateral and anterior PFC (t = 2.44, p < .02, uncorrected).

We also assessed the triple interaction between dAI covariance, IRI-EC, and EMP-SoVT to test whether EMP-SoVT modulates dAI networks significantly better than IRI-EC. A conjunction analysis of findings from Figure 5.2 and findings of a triple interaction (both at FEW, p < .05) revealed a stronger modulation of left dAI covariance networks by EMP-SoVT than IRI-EC to left ventrolateral prefrontal regions.

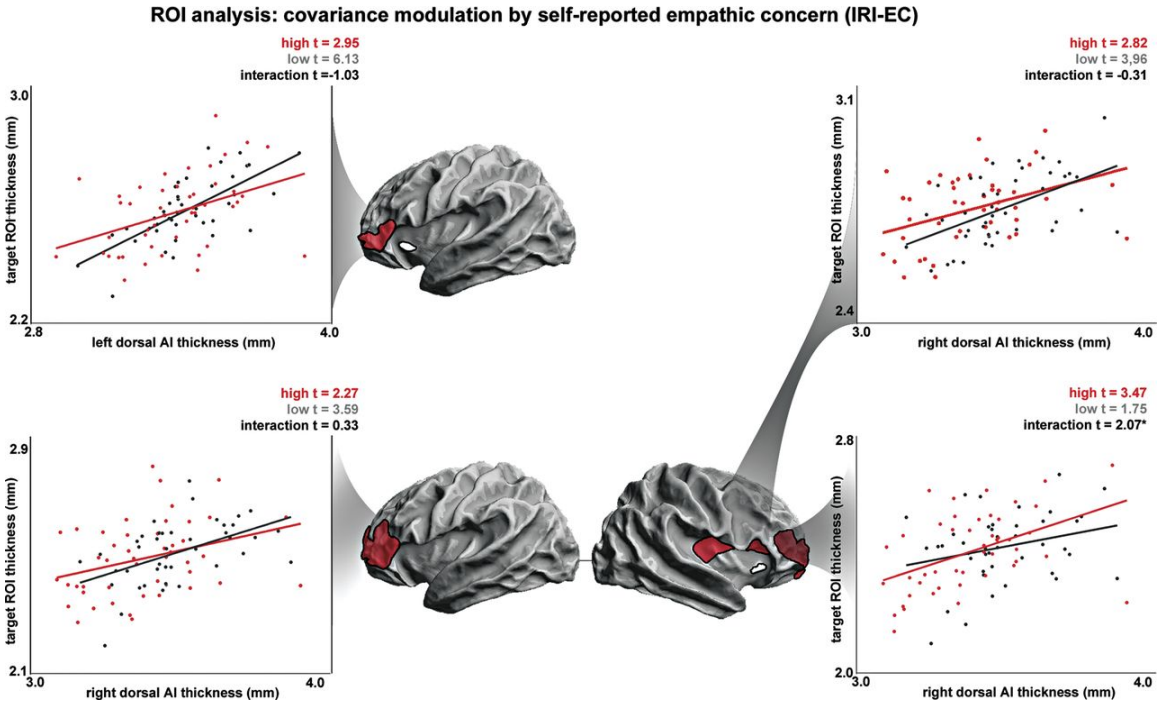


Figure 5.3. Interactions between the degree of structural covariance of right dAI and individual differences in self-reported trait empathy using the Interpersonal Reactivity Index, IRI (Davis, 1983). Target regions were independently chosen based on the whole-brain findings of covariance modulation by EMP-SoVT (Figure 5.2). To illustrate the parametric interaction effect, regression fits are also shown after the group has been split into 2 groups with n = 47 subjects according to the median of the IRI-EC scale (i.e., IRI-EC = 20).

Lack of Simple Relationship Between Thickness and Differences in Empathic Responding

Analyzing simple correlations between individual differences in empathic responding and thickness of our seed regions did not indicate any correlations between brain structure and interindividual differences in EMP-SoVT (t < 1.64, p > .1, uncorrected) nor IRI-EC (t < 1.60 p > .1, uncorrected). In our data, simple regression analysis, thus, did not show a sufficient sensitivity to detect structural markers of individual differences in empathic responding within the dAI and aMCC themselves.

Discussion

Based on covariance analysis of MRI-based cortical thickness measurements, we assessed the relationship between structural brain networks of dAI and aMCC and individual differences in empathic responding in 94 females. We observed widespread structural correlations of dAI and aMCC to multiple frontolimbic, temporal, and midline regions. This finding extends previous anatomical and functional connectivity results to the domain of inter-regional structural covariance patterns in the brain, suggesting that these regions reflect highly integrated hubs in the brain (Mesulam & Mufson, 1982; Margulies et al., 2007; Deen et al., 2011). Importantly, structural covariance of bilateral dAI but not aMCC to frontolimbic network components was related to individual differences in average empathy state ratings during a video-based task depicting people in distress (Klimecki et al., 2012) and, to a lesser extent, also to empathic concern trait scores from the IRI self-report questionnaire (Davis, 1983). These results could not be accounted for by individual differences observed in negative affect ratings when subjects were exposed to distressing situations, suggesting that the observed modulation in structural covariance was specific to individual differences in empathic sensitivity of the subjects. Modulations were robust even after statistically correcting for global mean thickness in the same statistical model, indicating the observed network modulation was specific to these frontolimbic networks, and not driven by a general and diffuse global effect. Our findings, thus, provide first evidence for a contribution of specific structural frontolimbic brain networks to individual differences in social processing above and beyond individual differences in general affective reactivity.

MRI covariance analysis has been proposed to map structural networks *in vivo* that likely reflect the manifestations of persistent functional-trophic cross-talk between different brain networks, together with common genetic, developmental and pathological influences (Bernhardt et al., 2011; Bullmore & Sporns, 2009; Lerch et al., 2006; Raznahan et al., 2011; Zielinski et al., 2010). In the present study, structural covariance networks centered on dAI and aMCC were widespread and resembled findings from tract tracing in nonhuman primates, showing insula interconnections, especially of its anterior parts, with multiple prefronal, temporolimbic, and medial and lateral parietal regions (Amaral & Price, 1984; Augustine, 1996; Mesulam & Mufson, 1982; Morecraft, Geula, & Mesulam, 1992; Mufson & Mesulam, 1982), together with similarly extensive interconnections of aMCC (Pandya et al., 1981; Vogt & Pandya, 1987). Moreover, covariance networks of dAI and aMCC in our participants resembled findings from previous

resting-state functional connectivity analyses that studied connectional fingerprints of regions proximal to our seeds (Cauda et al., 2011; Margulies et al., 2007; Hohmann et al., 2012). These findings, thus, lend evidence from structural covariance analysis to the notion that the AI and aMCC are important hub regions within multiple, distributed cortico-cortical networks, a role suitable to integrate several different functional processes hypothesized to interact in social cognition (Bernhardt & Singer, 2012; Chang et al., 2012; Craig, 2009; Devinsky, Morrell, & Vogt, 1995; Hohmann et al., 2012; Kurth, Zilles, Fox, Laird, & Eickhoff, 2010; Shackman et al., 2011; Singer et al., 2009).

Our main goal was to relate structural covariance networks centered on dAI and aMCC to individual differences in empathic responding. In our previous functional MRI study using the same dynamic Socio-affective Video Task, we showed that activations in dAI and aMCC correlated with empathy ratings for high emotion videos (Klimecki et al., 2012). The current study extended these findings by showing that inter-regional covariance patterns of brain structures are modulated by individual differences in empathy. This was accomplished by showing parametric modulatory effects of state and trait variables on group-level covariance networks. In fact, we were able to show that individual differences in EMP-SoVT parametrically related to the degree of structural covariance between bilateral AI, but not aMCC, and network components in frontolimbic cortices. Patterns of findings were similar when controlling for overall mean thickness in the same statistical model, suggesting that the modulation of dAI covariance was relatively specific to the observed frontolimbic network components and not driven by a general and diffuse modulation of dorsal AI covariance to all brain regions.

Importantly, these findings were specific to empathy, and could not be accounted for by differences in reported negative affect when exposed to the very same distressing scenes, as shown by our additional analysis that controlled for NEG-SoVT in the same statistical model. In other words, our current data revealed a differentiation at the level of inter-regional structural covariance networks between individual differences in empathy and more general negative affectivity; such a difference was not visible when focusing on functions of the dAI and aMCC and their relation to empathic and negative affective subjective experience alone (Klimecki et al., 2012). This differentiation at the level of inter-regional covariance is noteworthy because we did not observe any simple correlations in our seed regions with interindividual differences in empathy. Our findings, thus, did not confirm a recent voxel-based morphometric analysis that has shown correlations between IRI and gray matter estimates (Banissy et al., 2012). These findings may have been driven by differences in study groups. Indeed, while Banissy et al.

(2012) studied a mixed sample of males and females, the current study was limited to assessing females only. Alternatively, given that inference in voxel-based morphometry is based on a relatively complex measure that may be driven by differences in cortical thickness, but also cortical folding, and possibly sources of misregistration and partial voluming (Ashburner & Friston, 2001; Bookstein, 2001; Winkler et al., 2010), our cortical thickness analysis may have offered a somewhat different assessment of local brain structure. Ultimately, our findings suggest that inter-regional network analyses—even in the structural domain—open a new avenue to better understand and possibly dissociate neural signatures of closely related socio-affective processes.

Dorsal AI networks that showed a covariance modulation by EMP-SoVT encompassed subregions of ventrolateral PFC together with lateral and medial anterior PFC, as well as limbic mid- and posterior insular cortex. Modulations were more restricted and only observed ipsilaterally for left dAI and bilaterally and more extended for right dAI, particularly to medial PFC and posterior insular regions. Whether an increased structural covariance between 2 given regions indicates the direct strengthening of a "physical" interconnection is unclear. Nevertheless, covariance analysis might be a particularly suitable technique to detect structural manifestations of persistent functional and trophic cross-talk between different brain networks and their relationship to individual differences in empathy. Interestingly, previous functional studies have indeed shown an important role of the observed target regions in social cognition processes relevant for empathy, such as emotion and action observation (Caspers, Zilles, Laird, & Eickhoff, 2010; Molenberghs, Cunnington, & Mattingley, 2012; Rizzolatti, Fogassi, & Gallese, 2001), perspective taking (Amodio & Frith, 2006; Bzdok et al., 2012; Van Overwalle, 2009), and emotional awareness (Craig, 2002, 2009). Indeed, ventrolateral PFC subregions, such as BA44 and BA45, together with more anterior BA47 and BA10 participate in the observation of emotional facial expressions (Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003; Hennenlotter et al., 2005; Jabbi, Swart, & Keysers, 2007; Leslie, Johnson-Frey, & Grafton, 2004; Wicker et al., 2003). BA44, a region also thought to play an important role in action observation (Caspers et al., 2010; Gazzola & Keysers, 2009; Molenberghs et al., 2012; Rizzolatti et al., 2001), is also preferentially activated in empathy paradigms that use pictorial stimuli depicting others in pain (Lamm et al., 2011). Notably, previous functional MRI connectivity analysis showed increased functional coupling between fronto-insular and BA 45 subregions during the observation of emotional facial expressions, suggesting an influence of such fronto-insular network interactions on empathy (Jabbi & Keysers, 2008). Functional connectivity data have also suggested a

specific interaction during the observation of pain in others between dAI and more anterior dorsomedial PFC (Zaki et al., 2007), one of the most consistently activated regions during cognitive perspective taking (Amodio & Frith, 2006; Bzdok et al., 2012; Van Overwalle, 2009). Last, our finding of increased covariance between right dAI and mid-posterior insula in high empathizers may support models that relate intra-insular connectivity to empathy, interoception, and subjective awareness (Craig, 2002, 2009). Altogether, our findings of dAI covariance increases to the aforementioned regions suggests that high empathizers co-engage dAI and networks involved in social cognition more frequently, possibly leading to common morphological growth patterns. Such a hypothesis may be directly tested in longitudinal studies, which assess changes in inter-regional structural covariance patterns in subjects undergoing socio-affective training, ideally over sufficiently long periods to detect structural plasticity. Moreover, these studies may reveal further insights on a possible inter-hemispheric difference of dAI networks (Craig, 2009). Such studies will ideally employ more explicit behavioral testing on processes related to action observation, perspective taking, and awareness than the current work.

We additionally tested for a modulation of structural covariance by IRI-EC scores. While we failed to observe whole brain findings, post-hoc analysis revealed IRI-EC modulations between dAI and independently generated regions of interest from the EMP-SoVT analysis. Indeed, this analysis revealed higher covariance between right dAI and anterior and ventrolateral PFC in subjects with higher self-reported trait empathic concern, although at a weaker level than state empathy. These results provide additional, questionnaire-based support that the structural integration of dAI to other networks relates to inter-individual differences in empathy.

Arguably, the lack of whole brain findings when using IRI-EC might also indicate that state empathy ratings within a computer-based task, such as EMP-SoVT, offer an increased sensitivity relative to self-report trait questionnaire measures, such as IRI. A triple interaction analysis between dAI covariance, IRI-EC, and EMP-SoVT directly confirmed this suggestion for particular AI networks, showing a more marked modulation of covariance between left dAI and left ventrolateral PFC by EMP-SoVT than by IRI-EC. A lower sensitivity of self-report questionnaire data may stem from an increased influence of factors related to self-image, social desirability, and ‘cold’ cognition effects. Conversely, experimental state ratings may more directly tap into immediate and ‘hot’ socio-affective processes and, thus, allow a more veridical assessment of empathy and its modulation. In fact, a previous meta-analysis of functional MRI

studies also observed more robust modulations of dAI and aMCC activity during empathy for pain in others by state as opposed to trait scores (Lamm et al., 2011).

No noteworthy modulations of aMCC covariance strength by individual differences in empathic responding were observed. These findings support the notion that dAI, and its connectivity patterns, might play a different role in empathy than aMCC. A role of dAI in empathy may stem from the contribution of anterior insular segments to the generation of current and predictive feeling states, together with certainty computations that ultimately facilitate decision making in socio-affective contexts (Singer et al., 2009). A previous functional pattern analysis furthermore suggested that computations in dAI may more closely relate to prediction and sharing of affect, while aMCC computations rather reflect either nociceptive processing (Corradi-Dell'Acqua et al., 2011) or motor-related processes associated with pain (Morrison, Peelen, & Downing, 2007). Given the cytoarchitectonic, connectional, and functional heterogeneity of cingulate and insular cortex (Deen, et al., 2011; Gallay, Gallay, Jeanmonod, Rouiller, & Morel, 2012; Kurth, et al., 2010; Margulies, et al., 2007; Shackman, et al., 2011; Vogt, Vogt, Farber, & Bush, 2005), future studies are needed to provide a more comprehensive mapping of these subregions and their relationships to empathy and affect. In particular, a more detailed investigation of differences between dAI and more ventral aspects of the anterior insula may be of interest, as this subregion is frequently involved in general emotional processing and robustly coactivated with the amygdala (Mutschler et al., 2009; Kurth et al., 2010).

Our study adds structural MRI evidence to the understanding of individual differences in empathic responding, building on previous functional MRI literature. Indeed, several functional studies have shown a relationship between empathy-related brain responses, particularly of AI, and individual differences of empathy as assessed via trait-questionnaires such as the IRI, trial-by-trial affective ratings during the functional task, or subjective state measures of empathic concern assessed just after scanning (Singer et al., 2004; Singer et al., 2006; Hein et al., 2010; Lamm et al., 2011). On the other hand, we did not address the modulation of empathy by contextual factors, such as perceived fairness (Singer et al. 2006), appraisal (Lamm, Batson, & Decety, 2007; Lamm, Nusbaum, Meltzoff, & Decety, 2007; Hein & Singer, 2008; Engen & Singer, 2012), or perceived group membership (Hein et al., 2010). In this light, future work is needed that brings together structural brain network data, individual differences, and well as the flexibility of empathic responding in different contexts.

In summary, our findings show a contribution of inter-regional structural covariance networks in frontolimbic cortical regions to females' individual differences in empathic responding above and beyond negative affect. Thus, women with a high propensity to react empathically to the distress of others may have a stronger structural integration of dAI to other frontolimbic networks known to participate in social cognition. Ultimately, our structural covariance findings open a new methodological window to assess the complexity of brain networks underlying emotional and social processes.

Funding

This work was supported by the University of Zurich ("Neural Foundations of Empathy Learning," Forschungskredit, grant number 53990601 to O.M.K.), grants to T.S. from the Neuroscience Center Zurich, the Swiss National Science Foundation (SNF NCCR Affective Sciences: "Neuronal and developmental mechanisms of human empathy"; SNF NCCR Neural Plasticity and Repair: "Plasticity of the empathic and attentional brain"), the Foundation for Research in Science and Humanities at the University of Zurich (Neural foundations of empathy training). The research leading to these results has also received funding from the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013)/ERC Grant agreement no 205557 [EMPATHICBRAIN].

Notes

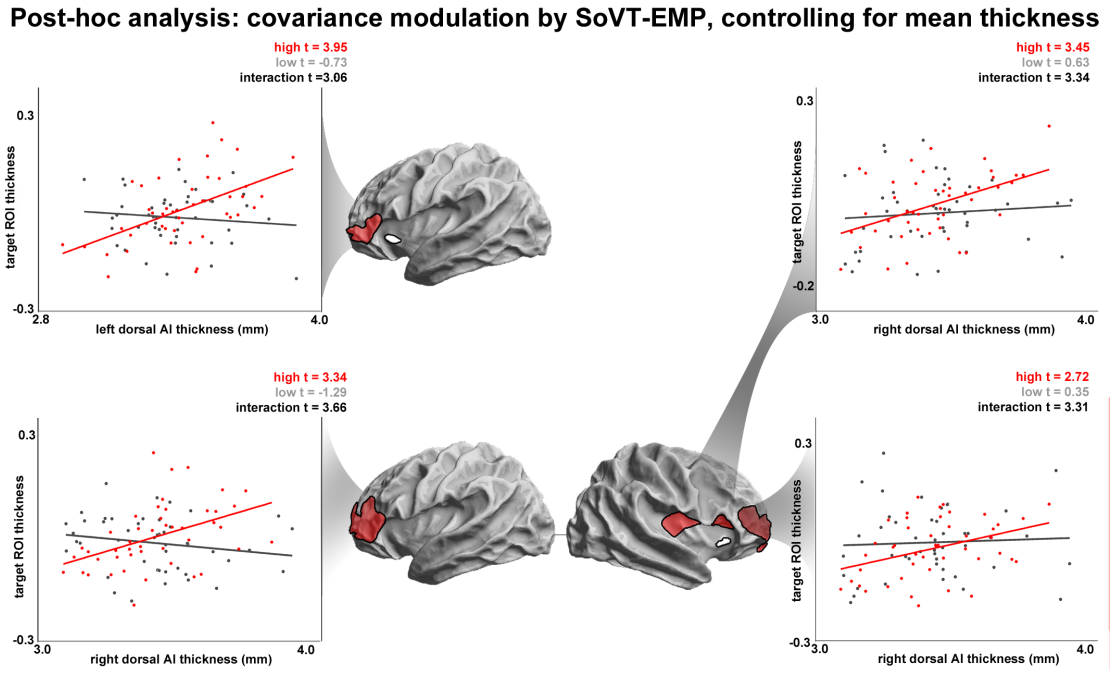
We thank all participants involved in this study. We are grateful to Nikolaus Steinbeis for insightful comments on the manuscript and Lisa Wittenburg for carefully proofreading the final version of this article.

Conflict of Interest: None declared.

Footnotes

- B.C.B and O.M.K. contributed equally to this work.

Supplementary Material



Supplementary Figure 5.1. Interactions between the degree of structural covariance of right dAI and individual differences in EMP-SoVT, controlling for global mean thickness in the same model. To illustrate the parametric interaction effect, regression fits are also shown after the group has been split into two groups with n = 47 subjects according to the median of EMP-SoVT (*i.e.* EMP-SoVT = 4.16).

6. General Discussion

Summary

The research described in this thesis followed two principal goals: firstly, to investigate neural, affective, and behavioral plasticity in the social brain and secondly, to dissociate the structural and functional neural basis of empathy from related emotions such as negative affect and compassion. Despite major advances in our understanding of how the training of motor, musical, and cognitive abilities shapes the brain (e.g., Draganski et al., 2004; Draganski et al., 2006; Karni et al., 1995; Olesen et al., 2004; Taubert et al., 2011), functional neural mechanisms underlying the learning of social emotions have not yet been unraveled. To answer this question, we studied functional neural plasticity induced by the training of empathy and compassion. Due to a lack of tasks allowing the reliable measurement of social affect and prosocial behavior in longitudinal designs, we first developed and validated two new paradigms: the Zurich Prosocial Game (ZPG) and the Socio-affective Video Task (SoVT). After validating these novel tasks in Study 1 and 2, respectively, we employed these tasks and found that: i) compassion training increased prosocial behavior in an implicit, training-unrelated task (Study 1), ii) training compassion augmented positive affect and reliably induced functional neural plasticity (Study 2), iii) training empathy and compassion induced different functional neural plasticity and distinct changes in self-reported affect (Study 3), and iv) structural covariance analysis allows the mapping of specific empathy-related networks (Study 4). The results obtained help to dissociate neural networks specific to empathy and compassion. In addition, these findings shed light on the functional plasticity involved in training social emotions. The observation that compassion training increased positive affect and prosocial behavior, suggests that compassion may be a new coping strategy that fosters resilience, while at the same time benefiting others. In light of the high prevalence rates of burnout in Western societies, training adaptive social emotions like compassion might be crucial to improving social skills, as well as mental and physical health.

Integrative Discussion

In Study 1, we introduced the ZPG, which allows repeated measurement of prosocial behavior in an ecologically valid setting. In addition, the ZPG can dissociate the influence of reciprocity, helping cost, and distress cues on helping behavior. Equipped with these properties, the ZPG offers a useful tool to assess changes in helping behavior in intervention studies. Using this novel task, we tested whether compassion was related to prosocial behavior. Our results revealed that training compassion, but not memory, augmented helping behavior, even though the task was unrelated to the training. Furthermore, inter-individual differences in practice duration predicted helping behavior under no-reciprocity conditions, emphasizing that practicing compassion particularly fosters altruistic behavior. Taken together, these findings provide first empirical evidence for the link between compassion training and helping behavior.

In Study 2, we first introduced the Socio-affective Video Task (SoVT) as a new paradigm for repeatedly studying social affect in the fMRI environment. The SoVT contains three matched video sets, which are composed of high emotion videos depicting people suffering and low emotion videos showing people in everyday activities. External validity of the SoVT was confirmed by showing that empathy state ratings in response to the videos correlated with empathic traits, as operationalized by a questionnaire (IRI, Davis, 1983). The matched nature of the three video sets was confirmed by the dependent variables valence, arousal, and empathy in a behavioral study with 265 participants. Moreover, an independent neuroimaging study with 94 participants confirmed the matched nature of the three video sets with respect to empathy, positive affect, and negative affect. On the neural level, the validation revealed that self-reported empathy for suffering during the SoVT engaged the AI and the aMCC—core regions of the empathy for pain network (Lamm et al., 2011, Fan et al., 2011). Moreover, empathy and negative affect in response to the social video stimuli were associated with overlapping neural activity patterns. Concordant with previously observed networks involved in processing negative social stimuli (e.g., Vrticka et al., 2011), seeing videos that depicted suffering augmented activations in the IFG, the thalamus and the temporal cortices. Interestingly, positive affect was related to distinct neural activations in the mOFC, which dovetails nicely with previous studies on positive affect in general (e.g., Kringelbach & Berridge, 2009).

In the past decade, major advances were made in deciphering the neural substrates of negative emotions, such as fear and anxiety (Delgado, Olsson, & Phelps, 2006). Similarly, the study of social emotions, like empathy, has begun to receive rising interest (Singer, 2012). However,

neuroscientific research on positive emotions is, surprisingly, still in its infancy. Functional neuroimaging studies of compassion, for instance, have so far been limited to very few cross-sectional experiments (e.g., Beauregard et al., 2009; J. W. Kim et al., 2009; Lutz et al., 2008). To reliably test how training compassion alters neural function, we performed multiple experiments, which are reported in Study 2. More precisely, using a short-term intervention study, we could provide first evidence that compassion training increased neural activations in the mOFC, the putamen, the pallidum and the VTA/SN. These findings were specific to compassion training, as they were not observed in a matched active control group. Moreover, the reliability of these findings was confirmed as the same neural pattern was observed in two independent compassion training studies (total *N* = 46) and a study with an expert meditation practitioner immersed in compassionate states. The observed compassion activations align with previous cross-sectional studies on compassion (Beauregard et al., 2009; J. W. Kim et al., 2009), as well as with cross-sectional neuroimaging studies of positive affect (e.g., Kringelbach & Berridge, 2009), affiliation (Strathearn et al., 2009; Vrticka, Andersson, Grandjean, Sander, & Vuilleumier, 2008), and maternal and romantic love (e.g., Bartels & Zeki, 2004). Astonishingly, the observed changes occurred even though participants were witnessing the suffering of others. Together with the increase of self-reported positive affect, which was also observed in response to other's affliction, these results suggest that resources strengthened through compassion training can be transferred to new situations and protect against potential feelings of distress. Interestingly, the observed increases in self-reported positive affect were not accompanied by decreases in negative affect. This unexpected finding probably indicates that, in contrast to traditional emotion regulation strategies like suppression or detachment (Gross, 2002), compassion does not rely on denial of the aversive content. On the contrary: the observed unique combination of a positive attitude with an awareness for negative events may be a crucial prerequisite for helping behavior, as seen in Study 1.

Study 3 was conducted to directly compare affective and functional neural plasticity of empathy and compassion training in the same participants. An active control group undergoing memory training controlled for retest effects. In this study, we observed that training empathy not only increased self-reports of empathy, but also negative affective experience. Remarkably, the increase of negative affect was not limited to situations that depicted other's suffering, but also extended to everyday life situations. This is conceptually similar to the process of hypersensitivity (Latremoliere & Woolf, 2009) and future studies could investigate how far the neural processes underlying these two phenomena resemble each other. On the neural level,

empathy training augmented activations in the DLPFC, the striatum, and a network that overlaps with cross sectional meta-analytic findings of empathy for pain and self-experienced pain in the AI and the aMCC (Lamm et al., 2011). Interestingly, the neural changes for HE and LE videos were very similar. This suggests that by training empathy, negative affect and corresponding neural substrates are augmented in response to both, everyday social situations and the witnessing of others' suffering. These results show that engaging in empathic resonance can increase aversive experiences, which in turn may be a risk factor for burnout. Encouraging changes were observed after subsequent compassion training, which increased positive affect and decreased negative affect compared to baseline, while not altering empathy levels. The combination of strong positive affect with normal negative affect levels found in Study 2 was replicated. This underlines the unique affective signature of compassion, which is distinct from emotion regulation strategies such as suppression or reappraisal (Gross, 2002). Future studies could explicitly compare compassion with traditional emotion regulation strategies on the neural and psychological level. In terms of functional brain changes induced through compassion training, we found augmented activations in the mOFC, the pregenual ACC and the ventral striatum. These findings show that even after a previous training in empathy, compassion training can augment activity in areas that were related to compassion in previous cross-sectional research (Beauregard et al., 2009; J. W. Kim et al., 2009) and in our own longitudinal research (Study 2). In addition, the present findings also align with results from studies on positive affect and reward in general (Haber & Knutson, 2010; Kringelbach & Berridge, 2009), as well as with results from studies examining affiliation (Strathearn et al., 2009; Vrticka et al., 2008) and love (Bartels & Zeki, 2004), in particular. Interestingly, empathy training and compassion training had different effects on activations in the cingulate cortex: while empathy training led to an increase of activation in aMCC, compassion training increased activity in the pACC. A recent meta-analysis of 939 studies (Shackman et al., 2011) found that the aMCC is crucial for processing negative affect and pain, and for cognitive control. Converging results were previously provided by a different meta-analysis (Beckmann, Johansen-Berg, & Rushworth, 2009), which reported that the aMCC is implicated in processing pain, conflict monitoring, and error detection. In addition, this part of the cingulate cortex was found to be highly connected to dorsal prefrontal regions (Beckmann et al., 2009). In line with this connectivity, training empathy in the current study increased activations in both the aMCC and the DLPFC. The comparison of cingulate cortex locations from fMRI studies on reward processing revealed a more anterior activation (Beckmann et al., 2009), which converges with the present observation of pACC involvement in compassion. Consistent with this notion, this

part of the cingulate cortex has been shown to be highly connected with the ventral striatum and the OFC (Beckmann et al., 2009), which dovetails nicely with augmented activation found in these structures after compassion training in Study 3.

In providing evidence for non-overlapping functional plasticity through training empathy and compassion, the present results extend neural plasticity research in humans. So far, neuroscientific plasticity studies have mainly focused on identifying mechanisms underlying the malleability of cognitive (Draganski et al., 2006; Olesen et al., 2004), musical (Jäncke, 2009; Taubert et al., 2011) or motor skills (Draganski et al., 2004; Karni et al., 1995). Meanwhile, research on affective plasticity has been quite limited, although some studies have investigated negative emotions, especially in the context of fear conditioning and fear extinction (for recent review, see Milad & Quirk, 2012). Given the scarcity of plasticity research on positive emotions, the observed malleability of social affect is of upmost importance for the implementation of training programs that increase resilience and teach adaptive responses to everyday distress and suffering.

Finally, Study 4 investigated empathy-related structural neural connectivity patterns in the participants of the main experiment in Study 2. To this end, we used seed regions in the AI and the aMCC, which were derived from a recent meta-analysis on empathy for pain (Lamm et al., 2011). Employing structural covariance analysis in order to delineate neural networks involved in socio-affective processes represents a novel approach. In a first step, we confirmed that structural covariance patterns of the AI and the aMCC resemble results from functional resting-state connectivity studies with seeds in equivalent regions (Cauda et al., 2011; Margulies et al., 2007). Furthermore, our data replicate tract-tracing results of insula connectivity in human and non-human primates (Augustine, 1996). In a next step, we showed that structural coupling of the AI with regions in the prefrontal cortex and the posterior insula was specifically modulated by empathic responses to the SoVT. This finding could not be accounted for by negative affect. Furthermore, AI covariance with fronto-limbic networks was replicated for empathic concern, as measured by a questionnaire (IRI, Davis, 1983), albeit at a weaker statistical level. These findings emphasize that structural covariance analysis can help to dissociate specific neural patterns in the domain of empathy. This adds to the results from Study 2 where no difference was observed between functional activations related to empathy and negative affect. The findings from Study 4 also extend previous research on functional connectivity, which mapped differential AI and aMCC connectivity for self-experienced pain and empathy for pain (Zaki et al., 2007). The AI and the aMCC may, thus, play different roles in empathic processes: while the AI

might be engaged in representing feeling states and predicting social situations (Singer, Critchley, & Preuschoff, 2009), the aMCC may subserve the processing of nociceptive computations (Beckmann et al., 2009; Shackman et al., 2011). Finally, the results of study 4 support the use of "hot" online situational measures, as opposed to "cold" questionnaire measures in social neuroscience research (Lamm et al., 2011) that may be biased by social desirability.

Limitations and outlook

This paragraph discusses the methodological advantages and disadvantages of the neuroimaging methods employed and outlines potential future research directions related to the study of empathy and compassion in general.

Using functional MRI and two newly developed tasks, the training research presented here mapped behavioral, affective, and neural plasticity induced by short-term training of social emotions. On the neural level, the reported findings elucidated which brain areas were implicated in the early phases of training empathy and compassion. However, it remains open as to which neuronal mechanisms underlie the observed plasticity effects. The observed activation increases may have, for instance, resulted from dendritic growth or arborization, synaptic modifications, or axonal changes. In addition, it is still unresolved which neurotransmitter systems are involved in fostering empathy and compassion. In light of the observation that compassion training activated neural structures related to dopamine, oxytocin, and endogenous opioids, it would be extremely helpful to dissociate the neurochemical processes implicated in compassion. This could be accomplished in future studies by using specific ligands in positron emission tomography (PET). As PET is an invasive technique, its use should be limited to carefully designed studies addressing specific questions. Alternatively, the influence of certain neurochemical processes could be mapped by using pharmacological intervention studies. It would, for instance, be interesting to explore the neural activation pathways through which oxytocin facilitates compassionate states (Rockliff et al., 2011).

Using fMRI allowed us to measure activation changes in both, cortical and subcortical structures. As the achieved spatial resolution in the present data was too coarse to disentangle activations in specific nuclei, like the VTA and the SN, future studies employing more targeted sequences and MRI scans carried out at higher magnetic field strengths (e.g., at 7 Tesla) may improve the

differentiation of small neural structures. Furthermore, the susceptibility of the brain stem and adjacent structures to respiration- and pulsation-related artifacts could be taken into consideration in future studies. The influence of physiological noise on the BOLD signal might, for instance, be reduced by using cardiac gating or post-acquisition corrections related to physiological parameters (Guimaraes et al., 1998).

As the presented training studies only mapped functional neural changes that occurred over several days, it would be interesting to gain a deeper understanding of the temporal properties of neural plasticity in longer training studies. Here, several levels could be bridged, including functional plasticity and structural changes in grey matter, white matter, and fiber tracts, as well as changes in functional and structural connectivity. Paralleling recent research in the domain of motor skill acquisition (Taubert et al., 2011), it would also be interesting to test how the changes on these different levels relate to one another over time.

In order to test how empathic states and traits modulate structural relations between distal brain regions *in vivo*, Study 4 mapped the covariance of cortical thickness with socio-affective measures. Although it has not yet been established whether structural covariance indeed reflects structural connectivity, accumulating evidence suggests that the observed patterns reflect common trophic, genetic, and developmental influences (Zielinski et al., 2010). The external validity of structural covariance was further confirmed by the convergence with previous functional connectivity patterns (e.g., Kelly et al., 2012; Seeley et al., 2009). Finally, the structural covariance analysis presented here emphasizes the usefulness of studying neural networks in order to gain increased specificity. A variety of existing techniques, ranging from dynamic causal modeling (Friston, Harrison, & Penny, 2003) to psycho-physiological interactions (Friston et al., 1997) and graph theoretical analyses (Bullmore & Sporns, 2009), offer promising avenues for more connectivity-based investigations.

While the current results represent important first steps in the investigation of neural plasticity of social affect, more extensive longitudinal studies are needed to dissect the elementary processes involved in the training of empathy and compassion. In addition, future research could investigate how interpersonal differences relate to the effectiveness of specific interventions. To improve external validity, laboratory measures might be complemented by experience sampling devices that can capture naturally occurring behavior. In the long run, more extensive research on the malleability of social affect will help to advance our basic understanding of learning in the socio-emotional domain. In addition, the obtained findings will

provide insights into how educational systems and employee development activities can be shaped in order to increase personal resources and resilience in children and adults alike.

References

Adelmann, P. K., & Zajonc, R. B. (1989). Facial efference and the experience of emotion. *Annual Review of Psychology, 40*, 249-280.

Akitsuki, Y., & Decety, J. (2009). Social context and perceived agency affects empathy for pain: An event-related fMRI investigation. *Neuroimage*, 47, 722-734.

Amaral, D. G., & Price, J. L. (1984). Amygdalo-cortical projections in the monkey (Macaca fascicularis). *Journal of Comparative Neurology, 230*(4), 465-496.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (Revised 4th ed.). Washington, DC: Author.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7*(4), 268-277.

Aron, A., Fisher, H., Mashek, D. J., Strong, G., Li, H., & Brown, L. L. (2005). Reward, motivation, and emotion systems associated with early-stage intense romantic love. *Journal of Neurophysiology, 94,* 327-337.

Ashburner, J., & Friston, K. J. (2001). Why Voxel-Based Morphometry Should Be Used. *Neuroimage*, 14(6), 1238-1243. doi: 10.1006/nimg.2001.0961

Augustine, J. R. (1996). Circuitry and functional aspects of the insular lobe in primates including humans. *Brain Research Reviews, 22*(3), 229-244. doi: S0165017396000112 [pii]

Avenanti, A., Bueti, D., Galati, G., & Aglioti, S. M. (2005). Transcranial magnetic stimulation highlights the sensorimotor side of empathy for pain. *Nature Neuroscience, 8*, 955-960.

Avenanti, A., Paluello, I. M., Bufalari, I., & Aglioti, S. M. (2006). Stimulus-driven modulation of motor-evoked potentials during observation of others' pain. *Neuroimage*, 32, 316-324.

Bagby, R. M., Parker, J. D., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale--I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research, 38*(1), 23-32.

Banissy, M. J., Kanai, R., Walsh, V., & Rees, G. (2012). Inter-individual differences in empathy are reflected in human brain structure. *Neuroimage, 62*, 2034–2039.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders, 34*, 163–175.

Bartels, A., & Zeki, S. (2000). The neural basis of romantic love. *Neuroreport, 11,* 3829-3834.

Bartels, A., & Zeki, S. (2004). The neural correlates of maternal and romantic love. *Neuroimage*, 21(3), 1155-1166.

Bartlett, M. Y., & DeSteno, D. (2006). Gratitude and prosocial behavior: Helping when it costs you. *Psychological Science, 17*, 319–325.

Batson, C. D. (1991). *The altruism question: Towards a social-psychological answer*. Mahwah, NJ: Erlbaum.

Batson, C. D. (2009a). These things called empathy: Eight related but distinct phenomena. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 3-15). Cambridge, MA: MIT Press.

Batson, C. D. (2009b). Empathy-induced altruistic motivation. In M. Mikulincer & P. R. Shaver (Eds.), *Prosocial motives, emotions, and behavior* (pp. 15-34). Washington, DC: American Psychological Association.

Batson, C. D., & Coke, J. S. (1981). Empathy: A source of altruistic motivation for helping? In: J. P. Rushton & R.M. Sorrentino (Eds.) *Altruism and helping behavior* (pp 167–187). Hillsdale, NJ: Erlbaum.

Batson, C. D., Duncan, B. D., Ackerman, P., Buckley, T., & Birch, K. (1981). Is empathic emotion a source of altruistic motivation? *Journal of Personality and Social Psychology, 40,* 290-302.

Batson, C. D., Eklund, J. H., Chermok, V. L., Hoyt, V., & Ortiz, B. G. (2007). An additional antecedent of empathic concern: Valuing the welfare of the person in need. *Journal of Personality and Social Psychology, 93*, 65–74.

Batson, C. D., Fultz, J., & Schoenrade, P. A. (1987). Distress and empathy: Two qualitatively distinct vicarious emotions with different motivational consequences. *Journal of Personality, 55,* 19.

Batson, C. D., O'Quin, K., Fultz, J., Mary, V. M., & Alice M., I. A. M. (1983). Influence of self-reported distress and empathy on egoistic versus altruistic motivation to help. *Journal of Personality and Social Psychology, 45,* 706-718.

Baumeister, R. F., & Vohs, K. D. (2007). *Encyclopedia of social psychology*. Thousand Oaks, CA: Sage.

Beauregard, M., Courtemanche, J., Paquette, V., & St-Pierre, E. L. (2009). The neural basis of unconditional love. *Psychiatry Research, 172,* 93-98.

Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. (1996). Comparison of Beck Depression Inventories – IA and II in psychiatric outpatients. *Journal of Personality Assessment*, 67, 588–597.

Beckmann, M., Johansen-Berg, H., & Rushworth, M. F. S. (2009). Connectivity-Based Parcellation of Human Cingulate Cortex and Its Relation to Functional Specialization. *The Journal of Neuroscience, 29*(4), 1175-1190. doi: 10.1523/jneurosci.3328-08.2009

Bernhardt, B. C., Bernasconi, N., Concha, L., & Bernasconi, A. (2010). Cortical thickness analysis in temporal lobe epilepsy: reproducibility and relation to outcome. *Neurology, 74*, 1776–1784.

Bernhardt, B. C., Chen, Z., He, Y., Evans, A. C., & Bernasconi, N. (2011). Graph-theoretical analysis reveals disrupted small-world organization of cortical thickness correlation networks in temporal lobe epilepsy. *Cerebral Cortex, 21*(9), 2147-2157. doi: 10.1093/cercor/bhq291

Bernhardt, B. C., & Singer, T. (2012). The neural basis of empathy. *Annual Review of Neuroscience, 35,*1-23.

Bernhardt, B. C., Worsley, K. J., Besson, P., Concha, L., Lerch, J. P., Evans, A. C., & Bernasconi, N. (2008). Mapping limbic network organization in temporal lobe epilepsy using morphometric correlations: insights on the relation between mesiotemporal connectivity and cortical atrophy. *Neuroimage*, *42*, 515–524.

Bird, G., Silani, G., Brindley, R., White, S., Frith, U., & Singer, T. (2010). Empathic brain responses in insula are modulated by levels of alexithymia but not autism. *Brain, 133*, 1515-1525.

Bohnet, I., & Frey, B. S. (1999). The sound of silence in prisoner's dilemma and dictator games. *Journal* of Economic *Behavior & Organization, 38*, 43–57.

Bolton, G. E., Katok, E., & Zwick, R. (1998). Dictator game giving: Rules of fairness versus acts of kindness. *International Journal of Game Theory, 27*, 269–299.

Bookstein, F. L. (2001). "Voxel-Based Morphometry" Should Not Be Used with Imperfectly Registered Images. *Neuroimage*, 14(6), 1454-1462. doi: 10.1006/nimg.2001.0770

Botvinick, M., Jha, A. P., Bylsma, L. M., Fabian, S. A., Solomon, P. E., & Prkachin, K. M. (2005). Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. *Neuroimage*, 25, 312-319.

Bower, G. H. (1970). Analysis of a mnemonic device. *American Scientist, 58*, 496-510.

Brett, M., Anton, J. L., Valabregue, R., & Poline, J. B. (2002). Region of interest analysis using an SPM toolbox. *Neuroimage*, 16(2), 1140-1141.

Bufalari, I., Aprile, T., Avenanti, A., Di Russo, F., & Aglioti, S. M. (2007). Empathy for pain and touch in the human somatosensory cortex. *Cerebral Cortex, 17*, 2553-2561.

Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews Neuroscience, 10*(3), 186-198. doi: nrn2575 [pii]10.1038/nrn2575

Burgdorf, J., & Panksepp, J. (2006). The neurobiology of positive emotions. *Neuroscience & Biobehavioral Reviews, 30, 173-187.*

Burnstein, E., Crandall, C., & Kitayama, S. (1994). Some neo-Darwinian decision rules for altruism: Weighing cues for inclusive fitness as a function of the biological importance of the decision. *Journal of Personality and Social Psychology, 67*, 773–789.

Bzdok, D., Schilbach, L., Vogeley, K., Schneider, K., Laird, A. R., Langner, R., & Eickhoff, S. B. (2012). Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of

mind, and empathy. *Brain Structure and Function*. doi: 10.1007/s00429-012-0380-y

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: a critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin,* 11, 401-423.

Camerer, C. F. (2003). *Behavioral game theory: experiments in strategic interaction*. Princeton: Princeton University Press.

Caprara, G. V., Steca, P., Zelli, A., & Capanna, C. (2005). A new scale for measuring adults' prosocialness. *European* Journal *of Psychological* Assessment*, 21*, 77–89.

Carlo, G., Eisenberg, N., Troyer, D., Switzer, G., & Speer, A. L. (1991). The altruistic personality: When is it apparent? *Journal of Personality and Social Psychology, 61*, 450–458.

Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences, 100*(9), 5497-5502.

Carson, J. W., Keefe, F. J., Lynch, T. R., Carson, K. M., Goli, V., Fras, A. M., & Thorp, S. R. (2005). Loving-kindness meditation for chronic low back pain: results from a pilot trial. *Journal of holistic nursing: official journal of the American Holistic Nurses' Association, 23*, 287-304. doi: 10.1177/0898010105277651

Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *Neuroimage*, 50(3), 1148-1167. doi: 10.1016/j.Neuroimage.2009.12.112

Cattaneo, L., Sandrini, M., & Schwarzbach, J. (2010). State-dependent TMS reveals a hierarchical representation of observed acts in the temporal, parietal, and premotor cortices. *Cerebral Cortex*, 20, 2252-8.

Cauda, F., D'Agata, F., Sacco, K., Duca, S., Geminiani, G., & Vercelli, A. (2011). Functional connectivity of the insula in the resting brain. *Neuroimage*, 55(1), 8-23. doi: 10.1016/j.Neuroimage.2010.11.049

Chang, L. J., Yarkoni, T., Khaw, M. W., & Sanfey, A. G. (2012). Decoding the role of the insula in

human cognition: functional parcellation and large-scale reverse inference. Cerebral Cortex, 23, 739–749.

Cheng, Y., Chen, C., Lin, C. P., Chou, K. H., & Decety, J. (2010). Love hurts: An fMRI study. *Neuroimage*, 51, 923-929.

Cheng, Y., Lin, C. P., Liu, H. L., Hsu, Y. Y., Lim, K. E., Hung, D. et al. (2007). Expertise modulates the perception of pain in others. *Current Biology, 17,* 1708-1713.

Cialdini, R. B., Brown, S. L., Lewis, B. P., Luce, C., & Neuberg, S. L. (1997). Reinterpreting the empathy-altruism relationship: When one into one equals oneness. *Journal of Personality and Social Psychology,* 73, 481–494.

Corradi-Dell'Acqua, C., Hofstetter, C., & Vuilleumier, P. (2011). Felt and Seen Pain Evoke the Same Local Patterns of Cortical Activity in Insular and Cingulate Cortex. *The Journal of Neuroscience, 31*(49), 17996-18006. doi: 10.1523/jneurosci.2686-11.2011

Craig, A. D. (2002). How do you feel? Interoception: The sense of the physiological condition of the body. *Nature Reviews Neuroscience*, *3*, 655-666.

Craig, A. D. (2009). How do you feel – now? The anterior insula and human awareness. *Nature Reviews Neuroscience*, *10*, 59-70.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage*, 9(2), 179-194. doi: 10.1006/nimg.1998.0395

Damasio, A. R. (1994). Descartes' error and the future of human life. *Scientific American, 271*, 144.

Danziger, N., Faillenot, I., & Peyron, R. (2009). Can we share a pain we never felt? Neural correlates of empathy in patients with congenital insensitivity to pain. *Neuron, 67*, 203-212.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalogue of Selected Documents in Psychology, 10*, 85.

Davis, M. H. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44* , 113-126.

Decety, J., & Jackson, P. L. (2006). A social-neuroscience perspective on empathy. *Current Directions in Psychological Science, 15*, 54-58.

Decety, J., Echols, S., & Correll, J. (2010). The blame game: The effect of responsibility and social stigma on empathy for pain. *Journal of Cognitive Neuroscience, 22*, 985-997.

Deen, B., Pitskel, N. B., & Pelphrey, K. A. (2011). Three systems of insular functional connectivity identified with cluster analysis. *Cerebral Cortex, 21*(7), 1498-1506. doi: 10.1093/cercor/bhq186

Delgado, M. R., Olsson, A., & Phelps, E. A. (2006). Extending animal models of fear conditioning to humans. *Biological Psychology, 73*(1), 39-48. doi: 10.1016/j.biopsycho.2006.01.006

Depue, R. A., & Morrone-Strupinsky, J. V. (2005). A neurobehavioral model of affiliative bonding: Implications for conceptualizing a human trait of affiliation. *Behavioral and Brain Sciences, 28,* 313-350.

de Vignemont, F., & Singer, T. (2006). The empathic brain: How, when and why? *Trends in Cognitive Sciences, 10,* 435-441.

Devinsky, O., Morrell, M. J., & Vogt, B. A. (1995). Contributions of anterior cingulate cortex to behaviour. *Brain, 118*, 279-306.

Dimberg, U., & Öhman, A. (1996). Behold the wrath: Psychophysiological responses to facial stimuli. *Motivation and Emotion, 20*, 149-182.

Dovidio, J. F., Piliavin, J. A., Gaertner, S. L., Schroeder, D. A., & Clark, R. D. (1991). The arousal-cost-reward model and the process of intervention: A review of the evidence. In: M. S. Clark, (Ed.), *Prosocial behavior (*pp. 86–118). Newbury Park, CA: Sage.

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Neuroplasticity: Changes in grey matter induced by training. [10.1038/427311a]. *Nature, 427*(6972), 311-312.

Draganski, B., Gaser, C., Kempermann, G., Kuhn, H. G., Winkler, J., Büchel, C., & May, A. (2006). Temporal and Spatial Dynamics of Brain Structure Changes during Extensive Learning. *The Journal of Neuroscience, 26*(23), 6314-6317. doi: 10.1523/jneurosci.4628-05.2006

Eckel, C. C., Grossman, P. J. (1996). Altruism in anonymous dictator games. *Games and Economic Behavior, 16*, 181–191.

Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual Review of Psychology, 51,* 665-697.

Eisenberg, N., & Fabes, R.A. (1998). Prosocial development. In: W. Damon & N. Eisenberg, (Eds.) *Handbook of child psychology: Vol. 3. Social, emotional, and personality development (*pp. 701–778.*)*. New York, NY: Wiley.

Eisenberg, N., & Miller, P. (1987). The relation of empathy to prosocial and related behaviors. *Psychological Bulletin, 101*, 91–119.

Eisenberg, N., Fabes, R. A., Miller, P. A., Fultz, J., Shell, R., Mathy, R. M. et al. (1989). Relation of sympathy and personal distress to prosocial behavior: A multimethod study. *Journal of Personality and Social Psychology*, *57,* 55-66.

Eisenberger, N. I., & Cole, S. W. (2012). Social neuroscience and health: Neuropsychological mechanisms linking social ties with physical health. *Nature Neuroscience*, 15, 669-674.

Engen, H. & Singer, T. (in press). Empathy circuits. *Current Opinion in Neurobiology.*

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology, 73*, 2608-2611.

Fahle, M. (2005). Perceptual learning: Specificity versus generalization. *Current Opinion in Neurobiology, 15*, 154–160.

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games Econ Behav, 54*, 293–315.

Fan, Y., & Han, S. (2008). Temporal dynamic of neural mechanisms involved in empathy for pain: an event-related brain potential study. *Neuropsychologia, 46*(1), 160-173. doi: 10.1016/j.neuropsychologia.2007.07.023

Fan, Y., Duncan, N. W., de Greck, M., & Northoff, G. (2011). Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. *Neuroscience & Biobehavioral Reviews, 35*(3), 903-911. doi: 10.1016/j.neubiorev.2010.10.009

Fehr, C., Sprecher, S., & Underwood, L. (2009). *The Science of Compassionate Love:Theory Research and Application*. Chichester: Wiley.

Fehr, E., & Schmidt, K. (2006). The Economics of Fairness, Reciprocity and Altruism - Experimental Evidence and New Theories. In: S. C. Kolm & J. M. Ythier, (Eds). *Reciprocity and Altruism (*pp. 615–691). Amsterdam: North-Holland.

Fields, H. L., Hjelmstad, G. O., Margolis, E. B., & Nicola, S.M. (2007). Ventral Tegmental Area Neurons in Learned Appetitive Behavior and Positive Reinforcement. *Annual Review of Neuroscience,* 30, 289-316.

Filbey, F. M., Claus, E., Audette, A. R., Niculescu, M., Banich, M. T., Tanabe, J., . . . Hutchison, K. E. (2007). Exposure to the Taste of Alcohol Elicits Activation of the Mesocorticolimbic Neurocircuitry. *Neuropsychopharmacology, 33*(6), 1391-1401.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics, 10*(2), 171-178.

Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences, 97*(20), 11050-11055. doi: 10.1073/pnas.200033797

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage*, 9(2), 195-207. doi: 10.1006/nimg.1998.0396

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior, 6*, 347–369.

Fowler, J., & Christakis, N. (2010). Cooperative behavior cascades in human social networks. *Proceedings of the National Academy of Sciences, 107*, 5334–5338.

Fredrickson, B. L., Cohn, M. A., Coffey, K. A., Pek, J., & Finkel, S. M. (2008). Open hearts build lives: Positive emotions, induced through loving-kindness meditation, build consequential personal resources. *Journal of Personality and Social Psychology, 95*, 1045-1062.

Fredrickson, B. L., Tugade, M. M., Waugh, C. E., & Larkin, G. R. (2003). What good are positive

emotions in crisis? A prospective study of resilience and emotions following the terrorist attacks on the United States on September 11th, 2001. *Journal of Personality and Social Psychology*, 84, 365-376.

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and Modulatory Interactions in Neuroimaging. *Neuroimage*, 6(3), 218-229. doi: 10.1006/nimg.1997.0291

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *Neuroimage*, 19(4), 1273-1302. doi: 10.1016/s1053-8119(03)00202-7

Friston, K. J., Worsley, K. J., Frackowiak, R. S. J., Mazziotta, J. C., & Evans, A. C. (1994). Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping, 1*(3), 210-220. doi: 10.1002/hbm.460010306

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*, 459-473.

Frith, C., & Frith, U. (2006). The neural basis of mentalizing. *Neuron, 50*, 531-534.

Frith, U., & Frith, C. (2010). The social brain: allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*(1537), 165-176. doi: 10.1098/rstb.2009.0160

Fydrich, T., Renneberg, B., Schmitz, B., & Wittchen, H.-U. (1997). Achse II: Persönlichkeitsstörungen.Interviewheft. *SKID-II. Strukturiertes Klinisches Interview für DSM-IV.* . Göttingen: Hogrefe.

Gallay, D. S., Gallay, M. N., Jeanmonod, D., Rouiller, E. M., & Morel, A. (2012). The insula of Reil revisited: multiarchitectonic organization in macaque monkeys. *Cerebral Cortex, 22*(1), 175-190. doi: 10.1093/cercor/bhr104

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119*, 593-609.

Gazzola, V., & Keysers, C. (2009). The observation and execution of actions share motor and somatosensory voxels in all tested subjects: single-subject analyses of unsmoothed fMRI data.

Cerebral Cortex, 19(6), 1239-1255. doi: 10.1093/cercor/bhn181

Gentili, R., Han, C.E., Schweighofer, N., & Papaxanthis, C. (2010). Motor learning without doing: Trial-by-trial improvement in motor performance during mental training. *Journal of Neurophysiology, 104*, 774–783.

Gilbert, P. (2007). *Psychotherapy and counseling for depression*. London: Sage.

Gilbert, P. (2010). *Compassion-focused therapy*. Abingdon, Oxford, UK: Routledge Chapman & Hall.

Gilbert, P., & Irons, C. (2005). Focused therapies and compassionate mind training for shame and self-attacking. In: P. Gilbert (Ed.) *Compassion: Conceptualisations, research and use in psychotherapy (*pp 263–325*)*. London, UK: Routledge.

Gilbert, P., & Procter, S. (2006). Compassionate mind training for people with high shame and self-criticism: Overview and pilot study of a group therapy approach. *Clinical Psychology & Psychotherapy, 13*, 353–379.

Goetz, J. L., Keltner, D., & Simon-Thomas, E. (2010). Compassion: An evolutionary analysis and empirical review. *Psychological Bulletin, 136*, 351-374.

Golan, O., Ashwin, E., Granader, Y., McClintock, S., Day, K., Leggett, V., & Baron-Cohen, S. (2010). Enhancing emotion recognition in children with autism spectrum conditions: an intervention using animated vehicles with real emotional faces. *Journal of Autism and Developmental Disorders, 40*, 269-279. doi: 10.1007/s10803-009-0862-9

Goldin, P. R., McRae, K., Ramel, W., & Gross, J. J. (2008). The Neural Bases of Emotion Regulation: Reappraisal and Suppression of Negative Emotion. *Biological Psychiatry, 63*(6), 577-586.

Gong, G., He, Y., Chen, Z. J., & Evans, A. C. (2012). Convergence and divergence of thickness correlations with diffusion connections across the human cerebral cortex. *Neuroimage*, 59(2), 1239-1248. doi: 10.1016/j.Neuroimage.2011.08.017

Gottfried, J. A., & Dolan, R. J. (2004). Human orbitofrontal cortex mediates extinction learning while accessing conditioned representations of value. *Nature Neuroscience, 7*(10), 1144-1152.

Gottman, J. M., & Levenson, R. W. (1985). A valid measure for obtaining self-report of affect. *Journal of Consulting and Clinical Psychology, 53*, 151-160.

Grabenhorst, F., & Rolls, E. T. (2011). Value, pleasure and choice in the ventral prefrontal cortex. *Trends in Cognitive Sciences, 15*(2), 56-67.

Greitemeyer, T. (2009). Effects of songs with prosocial lyrics on prosocial behavior: Further evidence and a mediating mechanism. *Personality and Social Psychology Bulletin,* 35, 1500–1511.

Grèzes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Human Brain Mapping, 12*, 1-19.

Gross, J. J. (2002). Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology, 39*(3), 281-291. doi: 10.1017/s0048577201393198

Gross, J. J., & Muñoz, R. F. (1995). Emotion Regulation and Mental Health. *Clinical Psychology: Science and Practice, 2*(2), 151-164.

Gu, X., & Han, S. (2007). Attention and reality constraints on the neural processes of empathy for pain. *Neuroimage*, 36, 256-267.

Guimaraes, A. R., Melcher, J. R., Talavage, T. M., Baker, J. R., Ledden, P., Rosen, B. R., . . . Weisskoff, R. M. (1998). Imaging subcortical auditory activity in humans. *Human Brain Mapping, 6*(1), 33-41. doi: 10.1002/(sici)1097-0193(1998)6:1<33::aid-hbm3>3.0.co;2-m

Haber, S. N., & Knutson, B. (2010). The Reward Circuit: Linking Primate Anatomy and Human Imaging. *Neuropsychopharmacology, 35*(1), 4-26.

Han, X., Jovicich, J., Salat, D., van der Kouwe, A., Quinn, B., Czanner, S., . . . Fischl, B. (2006). Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. *Neuroimage*, 32(1), 180-194. doi: 10.1016/j.Neuroimage.2006.02.051

Hari, R., Forss, N., Avikainen, S., Kirveskari, E., Salenius, S., & Rizzolatti, G. (1998). Activation of human primary motor cortex during action observation: A neuromagnetic study. *Proceedings of the National Academy of Sciences of the USA, 95*, 15061-15065.

Harrison, N. A., Singer, T., Rotshtein, P., Dolan, R. J., & Critchley, H. D. (2006). Pupillary contagion: Central mechanisms engaged in sadness processing. *Social Cognitive and Affective Neuroscience, 1*, 5-17.

Hatfield, E., Rapson, R. L., & Le, Y. L. (2009). Emotional contagion and empathy. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 19-30) Cambridge, MA: MIT.

Hein, G., Silani, G., Preuschoff, K., Batson, C. D., & Singer, T. (2010). Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron*, 68, 149-160.

Hein, G., & Singer, T. (2008). I feel how you feel but not always: the empathic brain and its modulation. *Current Opinion in Neurobiology, 18*, 153–158.

Hennenlotter, A., Schroeder, U., Erhard, P., Castrop, F., Haslinger, B., Stoecker, D., . . . Ceballos-Baumann, A. O. (2005). A common neural basis for receptive and expressive communication of pleasant facial affect. *Neuroimage*, 26(2), 581-591.

Hoffman, E., McCabe, K., Shachat, K., & Smith, V. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior, 7*, 346–380.

Hoffman, M. L. (2000). *Empathy and moral development.* Cambridge, UK: Cambridge University Press.

Hohmann, C.F., Beard, N.A., Kari-Kari, P., Jarvis, N., & Simmons, Q. (2012). Effects of brief stress exposure during early postnatal development in Balb/CByJ mice: II. Altered cortical morphology. *Developmental Psychobiology, 54*, 723–735.

Hutcherson, C. A., Seppala, E. M., & Gross, J. J. (2008). Loving-kindness meditation increases social connectedness. *Emotion, 8*, 720–724.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology, 3*, e79.

Ickes, W. (1993). Empathic accuracy. *Journal of Personality, 61*, 587-610.

Immordino-Yang, M. H., McColl, A., Damasio, H., & Damasio, A. . (2009). Neural correlates of admiration and compassion. *Proc Natl Acad Sci USA, 106*(19), 8021-8026.

Jabbi, M., Bastiaansen, J., & Keysers, C. (2008). A common anterior insula representation of disgust observation, experience and imagination shows divergent functional connectivity pathways. *PLoS One, 3*(8), e2939.

Jabbi, M., & Keysers, C. (2008). Inferior frontal gyrus activity triggers anterior insula response to emotional facial expressions. *Emotion, 8*(6), 775-780.

Jabbi, M., Swart, M., & Keysers, C. (2007). Empathy for positive and negative emotions in the gustatory cortex. *Neuroimage*, 34(4), 1744-1753.

Jackson, P. L., Brunet, E., Meltzoff, A. N., & Decety, J. (2006). Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. *Neuropsychologia, 44*, 752-761.

Jackson, P. L., Meltzoff, A. N., & Decety, J. (2005). How do we perceive the pain of others? A window into the neural processes involved in empathy. *Neuroimage*, 24, 771-779.

Jäncke, L. (2005). *Methoden der Bildgebung in der Psychologie und den kognitiven Neurowissenschaften.* Stuttgart: Kohlhammer.

Jäncke, L. (2009). The plastic human brain. *Restorative Neurology and Neuroscience, 27*(5), 521-538. doi: 10.3233/rnn-2009-0519

Jezzard, P., Matthews, P. M., & Smith, S. M. (2001). *Functional MRI*. New York: Oxford University Press.

Kalisch, R. (2009). The functional neuroanatomy of reappraisal: Time matters. *Neuroscience & Biobehavioral Reviews, 33*(8), 1215-1226.

Karni, A., Meyer, G., Jezzard, P., Adams, M. M., Turner, R., & Ungerleider, L. G. (1995). Functional MRI evidence for adult motor cortex plasticity during motor skill learning. *Nature, 377*(6545), 155-158.

Kelly, C., Toro, R., Di Martino, A., Cox, C. L., Bellec, P., Castellanos, F. X., & Milham, M. P.

(2012). A convergent functional architecture of the insula emerges across imaging modalities. *Neuroimage*, 61(4), 1129-1142. doi: 10.1016/j.Neuroimage.2012.03.021

Keltner, D., & Goetz, J. L. (2007). Compassion. In R. F. Baumeister & K. D. Vohs (Eds.), *Encyclopedia of social psychology* (pp. 159-161). Thousand Oaks, CA: Sage Publications.

Keysers, C., & Gazzola, V. (2009). Expanding the mirror: vicarious activity for actions, emotions, and sensations. *Current Opinion in Neurobiology, 19*(6), 666-671. doi: 10.1016/j.conb.2009.10.006

Keysers, C., Wicker, B., Gazzola, V., Anton, J. L., Fogassi, L., & Gallese, V. (2004). A touching sight: SII/PV activation during the observation and experience of touch. *Neuron, 42*, 335-346.

Kim, J. S., Singh, V., Lee, J. K., Lerch, J., Ad-Dab'bagh, Y., MacDonald, D., . . . Evans, A. C. (2005). Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification. *Neuroimage*, 27(1), 210-221. doi: 10.1016/j.Neuroimage.2005.03.036

Kim, J. W., Kim, S.-E., Kim, J.-J., Jeong, B., Park, C.-H., Son, A. R., . . . Ki, S. W. (2009). Compassionate attitude towards others' suffering activates the mesolimbic neural system. *Neuropsychologia, 47*(10), 2073-2081. doi: 10.1016/j.neuropsychologia.2009.03.017

Klimecki, O. M., Leiberg, S., Lamm, C., & Singer, T. (2012). Functional Neural Plasticity and Associated Changes in Positive Affect After Compassion Training. *Cerebral Cortex*. doi: 10.1093/cercor/bhs142

Klimecki, O., & Singer, T. (2012). Empathic distress fatigue rather than compassion fatigue? Integrating findings from empathy research in psychology and social neuroscience. In B. Oakley, A. Knafo, G. Madhavan, & D. S. Wilson (Eds.), *Pathological altruism.* New York, NY: Oxford University Press.

Knutson, B., & Cooper, J. C. (2005). Functional magnetic resonance imaging of reward prediction. *Current Opinion in Neurology, 18,* 411-417.

Kober, H., Barrett, L. F., Joseph, J., Bliss-Moreau, E., Lindquist, K., & Wager, T. D. (2008). Functional grouping and cortical-subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *Neuroimage*, 42, 998-1031.

Kring, A. M., & Gordon, A. H. (1998). Sex differences in emotion: Expression, experience, and physiology. *Journal of Personality and Social Psychology, 74*, 686–703.

Kringelbach, M. L., & Berridge, K. C. (2009). Towards a functional neuroanatomy of pleasure and happiness. *Trends in Cognitive Sciences, 13*(11), 479-487.

Kuperberg, G. R., Broome, M. R., McGuire, P. K., David, A. S., Eddy, M., Ozawa, F., . . . Fischl, B. (2003). Regionally localized thinning of the cerebral cortex in schizophrenia. *Archives of General Psychiatry, 60*(9), 878-888. doi: 10.1001/archpsyc.60.9.878

Kurth, F., Zilles, K., Fox, P. T., Laird, A. R., & Eickhoff, S. B. (2010). A link between the systems: functional differentiation and integration within the human insula revealed by meta-analysis. *Brain Structure and Function, 214*(5-6), 519-534. doi: 10.1007/s00429-010-0255-z

Lamm, C., Batson, C., & Decety, J. (2007a). The neural substrate of human empathy: Effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience, 19,* 42-58.

Lamm, C., & Decety, J. (2008). Is the extrastriate body area (EBA) sensitive to the perception of pain in others? *Cerebral Cortex, 18*, 2369-2373.

Lamm, C., Decety, J., & Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage*, 54(3), 2492-2502.

Lamm, C., Nusbaum, H. C., Meltzoff, A. N., & Decety, J. (2007b). What are you feeling? Using functional magnetic resonance imaging to assess the modulation of sensory and affective responses during empathy for pain. *PLoS ONE, 2*, e1292.

Lamm, C., & Singer, T. (2010). The role of anterior insular cortex in social emotions. *Brain Structure and Function, 214*, 579-951.

Larsen, J. T., Hemenover, S. H., Norris, C. J., & Cacioppo, J. T. (2003). Turning adversity to advantage: On the virtues of the coactivation of positive and negative emotions. In L.G. Aspinwall, & U.M. Staudinger, (Eds.), *A psychology of human strengths: Perspectives on an emerging field* (pp.211-226). Washington (DC): American Psychological Association.

Larsen, J. T., McGraw, A. P., & Cacioppo, J.T. (2001). Can people feel happy and sad at the

same time? *Journal of Personality and Social Psychology*, 81, 684-696.

Latremoliere, A., & Woolf, C. J. (2009). Central Sensitization: A Generator of Pain Hypersensitivity by Central Neural Plasticity. *The Journal of Pain, 10*(9), 895-926. doi: 10.1016/j.jpain.2009.06.012

Leiberg, S., Klimecki, O., & Singer, T. (2011). Short-term compassion training increases prosocial behavior in a newly developed prosocial game. *PLoS One, 6*(3), e17798. doi: 10.1371/journal.pone.0017798

Lerch, J. P., & Evans, A. C. (2005). Cortical thickness analysis examined through power analysis and a population simulation. *Neuroimage*, 24(1), 163-173. doi: 10.1016/j.Neuroimage.2004.07.045

Lerch, J. P., Worsley, K., Shaw, W. P., Greenstein, D. K., Lenroot, R. K., Giedd, J., & Evans, A. C. (2006). Mapping anatomical correlations across cerebral cortex (MACACC) using cortical thickness from MRI. *Neuroimage*, 31(3), 993-1003. doi: 10.1016/j.Neuroimage.2006.01.042

Leslie, K. R., Johnson-Frey, S. H., & Grafton, S. T. (2004). Functional imaging of face and hand imitation: towards a motor theory of empathy. *Neuroimage*, 21(2), 601-607. doi: 10.1016/j.Neuroimage.2003.09.038

Levenson, R. W., & Ruef, A. M. (1992). Empathy: A physiological substrate. *Journal of Personality and Social Psychology, 63*, 234-246.

Levesque, J., Eugene, F., Joanette, Y., Paquette, V., Mensour, B., Beaudoin, G., . . . Beauregard, M. (2003). Neural circuitry underlying voluntary suppression of sadness. *Biological Psychiatry, 53*, 502-510.

Levine, R. V., Martinez, T. S., Brase, G., & Sorenson, K. (1994). Helping in 36 U.S. cities. *Journal of Personality and Social Psychology, 67*, 69–82.

Lipps, T. (1903). Einfühlung, innere Nachahmung, und Organempfindungen [Empathy, inner imitation, and sense-feelings]. *Archiv für die gesamte Psychologie, 1*, 185-204.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature, 412*(6843), 150-157.

Loup, F., Tribollet, E., Dubois-Dauphin, M., Pizzolato, G., & Dreifuss, J. J. (1989). Localization of oxytocin binding sites in the human brainstem and upper spinal cord: an autoradiographic study. *Brain Research, 500*(1-2), 223-230.

Lustig, C., Shah, P., Seidler, R., & Reuter-Lorenz, P. A. (2009), Aging, training, and the brain. *Neuropsychology Review, 19*, 504–522.

Lutz, A., Brefczynski-Lewis, J., Johnstone, T., & Davidson, R. J. (2008). Regulation of the neural circuitry of emotion by compassion meditation: Effects of meditative expertise. *PLoS ONE, 3,* e1897.

MacDonald, D., Kabani, N., Avis, D., & Evans, A. C. (2000). Automated 3-D Extraction of Inner and Outer Surfaces of Cerebral Cortex from MRI. *Neuroimage*, 12(3), 340-356. doi: 10.1006/nimg.1999.0534

Mansouri, F. A., Tanaka, K., & Buckley, M. J. (2009). Conflict-induced behavioural adjustment: a clue to the executive functions of the prefrontal cortex. *Nature Reviews Neuroscience*, 10, 141-152.

Margulies, D. S., Kelly, A. M., Uddin, L. Q., Biswal, B. B., Castellanos, F. X., & Milham, M. P. (2007). Mapping the functional connectivity of anterior cingulate cortex. *Neuroimage*, 37(2), 579-588. doi: S1053-8119(07)00409-0 [pii]10.1016/j.Neuroimage.2007.05.019

McCray, L. W., Cronholm, P. F., Bogner, H. R., Gallo, J. J., & Neill, R. A. (2008). Resident Physician Burnout: Is There Hope? *Family Medicine, 40*(9), 626-632.

McCullough, M., Worthington, E., & Rachal, K. (1997). Interpersonal forgiving in close relationships. *Journal of Personality and Social Psychology, 73*, 321–336.

Mechelli, A., Friston, K. J., Frackowiak, R. S., & Price, C. J. (2005). Structural Covariance in the Human Cortex. *The Journal of Neuroscience, 25*(36), 8303-8310. doi: 10.1523/jneurosci.0357-05.2005

Mehrabian, A. (1997). Relations among personality scales of aggression, violence, and empathy: Validational evidence bearing on the risk of eruptive violence scale. *Aggressive Behavior, 23*, 433-445.

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality, 40*, 525-543.

Mesulam, M. M., & Mufson, E. J. (1982). Insula of the old world monkey. III: Efferent cortical output and comments on function. *Journal of Comparative Neurology, 212*(1), 38-52. doi: 10.1002/cne.902120104

Milad, M. R., & Quirk, G. J. (2012). Fear Extinction as a Model for Translational Neuroscience: Ten Years of Progress. *Annual Review of Psychology, 63*(1), 129-151. doi: doi:10.1146/annurev.psych.121208.131631

Mitchell, J. P. (2009). Inferences about mental states. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 364*, 1309-1316.

Molenberghs, P., Cunnington, R., & Mattingley, J. B. (2012). Brain regions with mirror properties: a meta-analysis of 125 human fMRI studies. *Neuroscience & Biobehavioral Reviews, 36*(1), 341-349. doi: 10.1016/j.neubiorev.2011.07.004

Morecraft, R. J., Geula, C., & Mesulam, M. M. (1992). Cytoarchitecture and neural afferents of orbitofrontal cortex in the brain of the monkey. *Journal of Comparative Neurology*, 323(3), 341-358. doi: 10.1002/cne.903230304

Moriguchi, Y., Decety, J., Ohnishi, T., Maeda, M., Mori, T., Nemoto, K. et al. (2007). Empathy and judging other's pain: An fMRI study of alexithymia. *Cerebral Cortex, 17*, 2223-2234.

Morrison, I., Lloyd, D., di Pellegrino, G., & Roberts, N. (2004). Vicarious responses to pain in anterior cingulate cortex: Is empathy a multisensory issue? *Cognitive, Affective and Behavioral Neuroscience, 4,* 270-278.

Morrison, I., Peelen, M. V., & Downing, P. E. (2007). The sight of others' pain modulates motor processing in human cingulate cortex. *Cerebral Cortex, 17*(9), 2214-2222. doi: 10.1093/cercor/bhl129

Mufson, E. J., & Mesulam, M. M. (1982). Insula of the old world monkey. II: Afferent cortical input and comments on the claustrum. *Journal of Comparative Neurology, 212*(1), 23-37. doi: 10.1002/cne.902120103

Mutschler, I., Wieckhorst, B., Kowalevski, S., Derix, J., Wentlandt, J., Schulze-Bonhage, A., & Ball, T. (2009). Functional organization of the human anterior insular cortex. *Neuroscience Letters, 457*, 66–70.

Nelson, J. (2005). *Seeing through tears: Crying and attachment*. New York: Brunner-Routledge.

Nemiah, J. C., Freyberger, H., & Sifneos, P. E. (1976). Alexithymia: A view of the psychosomatic process. In O. W. Hill (Ed.), *Modern trends in psychosomatic medicine* (pp. 430-439). London, UK: Butterworths.

Neumann, R., & Strack, F. (2000). "Mood contagion": The automatic transfer of mood between persons. *Journal of Personality and Social Psychology, 79*, 211-223.

Nowak, M., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature, 437*, 1291–1298.

Nowak, M., Page, K., & Sigmund, K. (2000). Fairness versus reason in the ultimatum game. *Science, 289*, 1773–1775.

Ochsner, K. N., Bunge, S. A., Gross, J. J., & Gabrieli, J. D. (2002). Rethinking feelings: an FMRI study of the cognitive regulation of emotion. *Journal of Cognitive Neuroscience, 14*(8), 1215-1229.

Ogino, Y., Nemoto, H., Inui, K., Saito, S., Kakigi, R., & Goto, F. (2007). Inner experience of pain: Imagination of pain while viewing images showing painful events forms subjective pain representation in human brain. *Cerebral Cortex, 17*, 1139-1146.

Olesen, P. J., Westerberg, H., & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nature Neuroscience, 7*(1), 75-79.

Ostrowsky, K., Isnard, J., Ryvlin, P., Guénot, M., Fischer, C., & Mauguière, F. (2000). Functional mapping of the insular cortex: Clinical implication in temporal lobe epilepsy. *Epilepsia, 41*, 681-686.

Ostrowsky, K., Magnin, M., Ryvlin, P., Isnard, J., Guenot, M., & Mauguiere, F. (2002). Representation of pain and somatic sensation in the human insula: A study of responses to direct electrical cortical stimulation. *Cerebral Cortex, 12*, 376-385.

Pace, T. W. W., Negi, L. T., Adame, D. D., Cole, S. P., Sivilli, T. I., Brown, T. D., Issa, M. J., & Raison, C. L. (2009). Effect of compassion meditation on neuroendocrine, innate immune and behavioral responses to psychosocial stress. *Psychoneuroendocrinology,* 34, 87-98.

Pandya, D. N., Van Hoesen, G. W., & Mesulam, M. M. (1981). Efferent connections of the cingulate gyrus in the rhesus monkey. *Experimental Brain Research, 42*(3-4), 319-330.

Panksepp, J. (2011). Cross-Species Affective Neuroscience Decoding of the Primal Affective Experiences of Humans and Related Animals. *PloS one, 6*(9), e21236. doi: 10.1371/journal.pone.0021236

Pascual-Leone, A., Amedi, A., Fregni, F., & Merabet, L. (2005). The Plastic Human Brain Cortex. *Annual Review of Neuroscience, 28*(1), 377-401.

Phan, K. L., Fitzgerald, D. A., Nathan, P. J., Moore, G. J., Uhde, T. W., & Tancer, M. E. (2005). Neural substrates for voluntary suppression of negative affect: a functional magnetic resonance imaging study. *Biological Psychiatry, 57*(3), 210-219.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*, 515-526.

Preston, S. D., & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Science, 25*, 1-72.

Rainville, P. (2002). Brain mechanisms of pain affect and pain modulation. *Current Opinion in Neurobiology, 12*(2), 195-204.

Rangel, A., & Hare, T. (2010). Neural computations associated with goal-directed choice. *Current Opinion in Neurobiology, 20*(2), 262-270.

Raznahan, A., Lerch, Jason P., Lee, N., Greenstein, D., Wallace, G. L., Stockman, M., Clasen, L., Shaw P. W., & Giedd, J. N. (2011). Patterns of Coordinated Anatomical Change in Human Cortical Development: A Longitudinal Neuroimaging Study of Maturational Coupling. *Neuron, 72*(5), 873-884. doi: 10.1016/j.neuron.2011.09.028

Rizzolatti, G., & Craighero, L. (2004). The mirror neuron system. *Annual Review of Neuroscience, 27*, 169-192.

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature reviews Neuroscience, 11*, 264-274.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Brain Research: Cognitive Brain Research, 3*, 131-141.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature reviews Neuroscience, 2*(9), 661-670. doi: 10.1038/3509006035090060 [pii]

Rockliff, H., Gilbert, P., McEwan, K., Lightman, S., & Glover, D. (2008). Heart rate variability and salivary cortisol responses to compassion-focused imagery. *Journal of Clinical Neuropsychiatry*, 5, 132–139.

Rockliff, H., Karl, A., McEwan, K., Gilbert, J., Matos, M., & Gilbert, P. (2011). Effects of intranasal oxytocin on 'compassion focused imagery'. *Emotion, 11*(6), 1388-1396. doi: 10.1037/a0023861

Rosas, H. D., Liu, A. K., Hersch, S., Glessner, M., Ferrante, R. J., Salat, D. H., . . . Fischl, B. (2002). Regional and progressive thinning of the cortical ribbon in Huntington's disease. *Neurology, 58*(5), 695-701.

Saarela, M. V., Hlushchuk, Y., Williams, A. C., Schurmann, M., Kalso, E., & Hari, R. (2007). The compassionate brain: Humans detect intensity of pain from another's face. *Cerebral Cortex, 17,* 230-237.

Salzberg, S. (2002). *Loving-kindness: The revolutionary art of happiness*. Boston, MA: Shambhala.

Saxe, R. (2006). Why and how to study Theory of Mind with fMRI. *Brain Research, 1079*, 57-65.

Saxe, R., & Baron-Cohen, S. (2006). The neuroscience of theory of mind. *Social Neuroscience, 1*, i-ix.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in 'theory of mind'. *Neuroimage, 19*, 1835–1842.

Schernhammer, E. S., & Colditz, G. A. (2004). Suicide Rates Among Physicians: A Quantitative

and Gender Assessment (Meta-Analysis). *The American Journal of Psychiatry, 161*, 2295-2302. doi: 10.1176/appi.ajp.161.12.2295

Schiller, D., Levy, I., Niv, Y., LeDoux, J. E., & Phelps, E. A. (2008). From Fear to Safety and Back: Reversal of Fear in the Human Brain. *Journal of Neuroscience, 28*(45), 11517-11525.

Schultz, W. (2000). Multiple reward signals in the brain. *Nature reviews Neuroscience, 1,* 199-207.

Schultz, W. (2010). Subjective neuronal coding of reward: temporal value discounting and risk. *European Journal of Neuroscience*, 31, 2124-2135.

Seeley, W. W., Crawford, R. K., Zhou, J., Miller, B. L., & Greicius, M. D. (2009). Neurodegenerative Diseases Target Large-Scale Human Brain Networks. *Neuron, 62*(1), 42-52. doi: 10.1016/j.neuron.2009.03.024

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., Reiss, A. L., & Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience*, 27, 2349-56.

Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J., & Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nature Reviews Neuroscience, 12*(3), 154-167. doi: http://www.nature.com/nrn/journal/v12/n3/suppinfo/nrn2994_S1.html

Silani, G., Bird, G., Brindley, R., Singer, T., Frith, C., & Frith, U. (2008). Levels of emotional awareness and autism: An fMRI study. *Social Neuroscience, 3*, 97-112.

Simon-Thomas, E. R., Godzik, J., Castle, E., Antonenko, O., Ponz, A., Kogan, A., & Keltner, D. J. (2011). An fMRI study of caring vs self-focus during induced compassion and pride. Social *Cognitive* and Affective *Neuroscience* [E-pub ahead of print].

Singer, T. (2012). The past, present and future of social neuroscience: A European perspective. *Neuroimage*, *61*, 437-49. doi: 10.1016/j.Neuroimage.2012.01.109

Singer, T., Critchley, H. D., & Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. *Trends in Cognitive Sciences, 13,* 334-340.

Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *The Year in Cognitive Neuroscience 2009: Annals of the New York Academy of Sciences, 1156,* 81-96.

Singer, T., & Leiberg, S. (2009). Sharing the emotions of others: The neural bases of empathy. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences IV* (pp. 971-984). Cambridge, MA: MIT.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature, 439,* 466-469. doi: 10.1038/nature04271

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R., & Frith, C. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science, 303,* 1157-1162.

Singer, T., Snozzi, R., Bird, G., Petrovic, P., Silani, G., Heinrichs, M. et al. (2008). Effects of oxytocin and prosocial behavior on Brain Responses to direct and vicariously experienced pain. *Emotion, 8,* 781-791.

Singer, T., & Steinbeis, N. (2009). Differential roles of fairness- and compassion-based - motivations for cooperation, defection, and punishment. *Annals of the New York Academy of Sciences*,*1167*, 41–50.

Slaski, M. & Cartwright, S. (2003). Emotional intelligence training and its implications for stress, health and performance. *Stress Health, 19,* 233–239.

Sprecher, S., & Fehr, B. (2005). Compassionate love for close others and humanity. *Journal of Social and Personal Relationships, 22*(5), 629.

Squire, L. R., Stark, C. E. L., & Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience, 27*(1), 279-306.

Strathearn, L., Fonagy, P., Amico, J., & Montague, P. R. (2009). Adult Attachment Predicts Maternal Brain and Oxytocin Response to Infant Cues. *Neuropsychopharmacology, 34*(13), 2655-2666. doi: http://www.nature.com/npp/journal/v34/n13/suppinfo/npp2009103s1.html

Taubert, M., Lohmann, G., Margulies, D. S., Villringer, A., & Ragert, P. (2011). Long-term effects of motor training on resting-state networks and underlying brain structure. *Neuroimage*, 57(4), 1492-1498. doi: 10.1016/j.Neuroimage.2011.05.078

Trivers, R. (1971). The evolution of reciprocal altruism. *The Quarterly review of biology, 46*, 35–57.

Twenge, J. M., Baumeister, R. F., DeWall, C. N., Ciarocco, N. J., & Bartels, J. M. (2007). Social exclusion decreases prosocial behavior. *Journal of personality and social psychology*, 92, 56–66.

Valeriani, M., Betti, V., Le Pera, D., De Armas, L., Miliucci, R., Restuccia, D. et al. (2008). Seeing the pain of others while being in pain: A laser-evoked potentials study. *Neuroimage*, 40, 1419-1428.

Van Lange, P. A. M., Bekkers, R., Schuyt, T. N. M., & Vugt, M. V. (2007). From games to giving: Social value orientation predicts donations to noble causes. *Basic and Applied Social Psychology, 29*, 375–384.

Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Human Brain Mapping, 30*(3), 829-858. doi: 10.1002/hbm.20547

Verhaeghen, P., & Marcoen, A. (1996). On the mechanisms of plasticity in young and older adults after instruction in the method of loci: Evidence for an amplification model. *Psychology and Aging*, 11, 164–178.

Vogt, B. A. (2005). Pain and emotion interactions in subregions of the cingulate gyrus. *Nature reviews Neuroscience, 6*(7), 533-544. doi: 10.1038/nrn1704

Vogt, B. A., Vogt, L., Farber, N. B., & Bush, G. (2005). Architecture and neurocytology of monkey cingulate gyrus. *The Journal of Comparative Neurology, 485*(3), 218-239. doi: 10.1002/cne.20512

Vrticka, P., Andersson, F., Grandjean, D., Sander, D., & Vuilleumier, P. (2008). Individual Attachment Style Modulates Human Amygdala and Striatum Activation during Social Appraisal. *PLoS One, 3*(8), e2868.

Vrticka, P., Sander, D., & Vuilleumier, P. (2011). Effects of emotion regulation strategy on Brain Responses to the valence and social content of visual scenes. *Neuropsychologia, 49*(5), 1067-1082.

Walter, H., von Kalckreuth, A., Schardt, D., Stephan, A., Goschke, T., & Erk, S. (2009). The Temporal Dynamics of Voluntary Emotion Regulation. *PLoS One, 4*(8), e6726.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology, 54*, 1062-1070.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: the common neural basis of seeing and feeling disgust. *Neuron, 40*(3), 655-664.

Wilke, H. & Lanzetta, J. T. (1970). The obligation to help: The effects of amount of prior help on subsequent helping behavior. *Journal of experimental social psychology, 6*, 488–493.

Winkler, A. M., Sabuncu, M. R., Yeo, B. T. T., Fischl, B., Greve, D. N., Kochunov, P., . . . Glahn, D. C. (2012). Measuring and comparing brain cortical surface area and other areal quantities. *Neuroimage*, 61(4), 1428-1443. doi: 10.1016/j.Neuroimage.2012.03.026

Wittchen, H. U., Zaudig, M., & Fydrich, T. (1997). *Structured clinical interview for DSM-IV, german version*. Göttingen: Hogrefe.

Winkler, A. M., Kochunov, P., Blangero, J., Almasy, L., Zilles, K., Fox, P. T., Duggirala, R., & Glahn, D.C. (2010). Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *Neuroimage, 53*,1135–1146.

Worsley, K. J., Andermann, M., Koulis, T., MacDonald, D., & Evans, A. C. (1999). Detecting changes in nonisotropic images. *Human Brain Mapping, 8*(2-3), 98-101.

Worsley, K. J., Taylor, J. E., Carbonell, F., Chung, M. K., Duerden, E., Bernhardt, B. C., Lyttelton, O., Boucher, M., & Evans, A.C. (2009). SurfStat: A Matlab toolbox for the statistical analysis of univariate and multivariate surface and volumetric data using linear mixed effects models and random field theory. *Neuroimage, 47*, (Suppl 1): S102.

Xu, X., Zuo, X., Wang, X., & Han, S. (2009). Do you feel my pain? Racial group membership modulates empathic neural responses. *Journal of Neuroscience, 29*, 8525-8529.

Zaki, J., & Ochsner, K. (2012). The neuroscience of empathy: progress, pitfalls and promise. *Nature Neuroscience, 15*(5), 675-680.

Zaki, J., Ochsner, K. N., Hanelin, J., Wager, T. D., & Mackey, S. C. (2007). Different circuits for

different pain: patterns of functional connectivity reveal distinct networks for processing pain in self and others. *Social Neuroscience, 2*(3-4), 276-291. doi: 10.1080/17470910701401973

Zeki, S. (2007). The neurobiology of love. *FEBS Letters, 581,* 2575-2579.

Zielinski, B. A., Gennatas, E. D., Zhou, J., & Seeley, W. W. (2010). Network-level structural covariance in the developing brain. *Proceedings of the National Academy of Sciences, 107*(42), 18191-18196. doi: 10.1073/pnas.1003109107

PUBLICATIONS

Research Articles

Gamer, M., **Klimecki, O.**, Bauermann, T., Stoeter, P. & Vossel G. (2009). *fMRI-activation patterns in the detection of concealed information rely on memory-related effects*. Social, Cognitive and Affective Neuroscience.

Leiberg, S., **Klimecki, O.**, & Singer, T. (2011). *Short-term compassion training increases prosocial behavior in a newly developed prosocial game*. PLoS One, 6(3).

Klimecki, O., Leiberg, S., Lamm, C., Singer, T. (2012). *Functional neural plasticity and associated changes in positive affect after compassion training*. Cerebral Cortex. doi: 10.1093/cercor/bhs142

Klimecki, O., Leiberg, S., Ricard, M., Singer, T. (2013). Differential Pattern of Functional Brain Plasticity after Compassion and Empathy Training. Soc Cogn Affect Neurosci. doi: 10.1093/scan/nst060

Bernhardt, B.*, **Klimecki, O.***, Leiberg, S., Singer T. (2013). *Structural Covariance Networks of the Dorsal Anterior Insula Predict Females.* (*both authors contributed equally to this work). Cerebral Cortex. doi: 10.1093/cercor/bht072

Book Chapters

Klimecki, O., & Singer, T. (2012). Empathic distress fatigue rather than compassion fatigue? Integrating findings from empathy research in psychology and social neuroscience. In B. Oakley, A.Knafo, G. Madhavan, & D. S. Wilson (Eds.), *Pathological altruism* (pp. 368-383). New York: Oxford University Press.

Klimecki, O., & Singer, T. (2013). Empathy from the perspective of social neuroscience. In: J. Armony & P. Vuilleumier (Eds.), *Handbook of Human Affective Neuroscience.* New York: Cambridge University Press.

PRESENTATIONS

Klimecki, O., Leiberg, S., Lamm, C., & Singer, T. (2012). *Neural and behavioral changes related to compassion training*. Poster presented at the Alpine Brain Imaging Meeting, Champery, Switzerland.

Klimecki, O., & Singer, T. (2011). *Compassion vs. empathic distress and memory training as an active control group*. Talk presented at Workshop "How to Train Compassion", organized by Max-Planck Institute Leipzig, taking place in Berlin, Germany.

Klimecki, O., Leiberg, S., Lamm, C., & Singer, T. (2011). *Neural and behavioral changes related to compassion training*. Poster presented at the Workshop “The Social Brain", organized by MRC Cognition and Brain Sciences Unit, Cambridge, United Kingdom.

Leiberg, S., Clos, M., Sorger, B., **Klimecki, O.**, Goebel, R., & Singer, T. (2010). *Effects of real-time functional magnetic resonance imaging (rt-fMRI) neurofeedback from the ventral striatum in compassion meditation training.* Poster presented at the “Alpine Brain Imaging Meeting”, Champery, Switzerland.

Leiberg, S., Clos, M., Sorger, B., **Klimecki, O.**, Goebel, R., & Singer, T. (2010). *Effects of real-time functional magnetic resonance imaging (rt-fMRI) neurofeedback from the ventral striatum in compassion meditation training.* Poster presented at the Human Brain Mapping Conference, Barcelona, Spain.

Gamer, M., **Klimecki, O.**, & Vossel, G. (2009). *Emotionale Modulation des Gedächtnisses im Tatwissentest.* Poster presented at the 35. Arbeitstagung Psychophysiologie und Methodik, Leipzig, Germany.

Leiberg, S., **Klimecki, O.**, & Singer, T. (2008). *Training compassion with real-time Functional Magnetic Resonance Imaging - First Insights.* Poster presented at the International Symposium "Foundations of Human Social Behavior", University of Zurich, Switzerland

Gamer, M., **Klimecki, O.**, Bauermann, T., Stoeter, P., & Vossel, G. (2007). *Relevanz von Reaktionskonflikten und Gedächtnisprozessen bei der Verarbeitung verheimlichten Wissens.* Poster presented at the Psychologie & Gehirn, Dortmund, Germany.